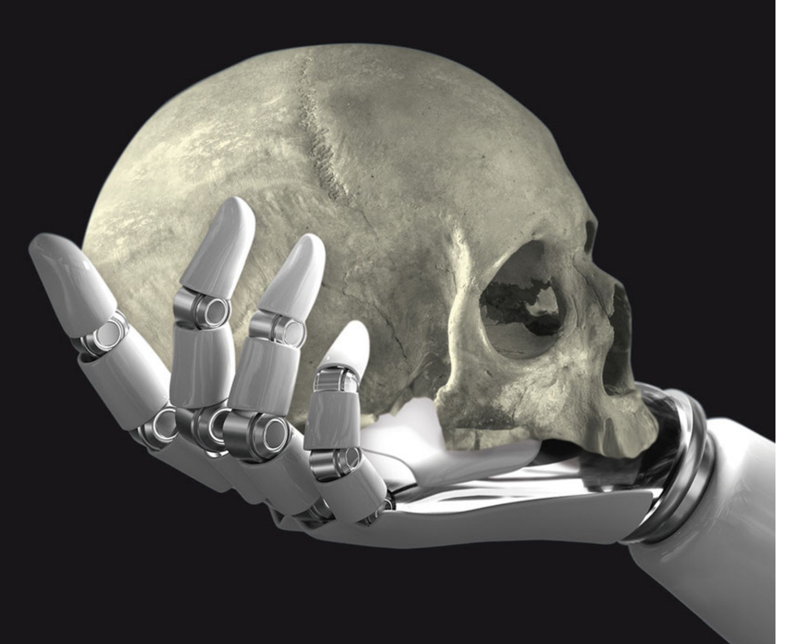
FERNANDO BONETE VIZCAÍNO

LA GUERRA IMAGINARIA

Desmontando el mito de la inteligencia artificial con Asimov

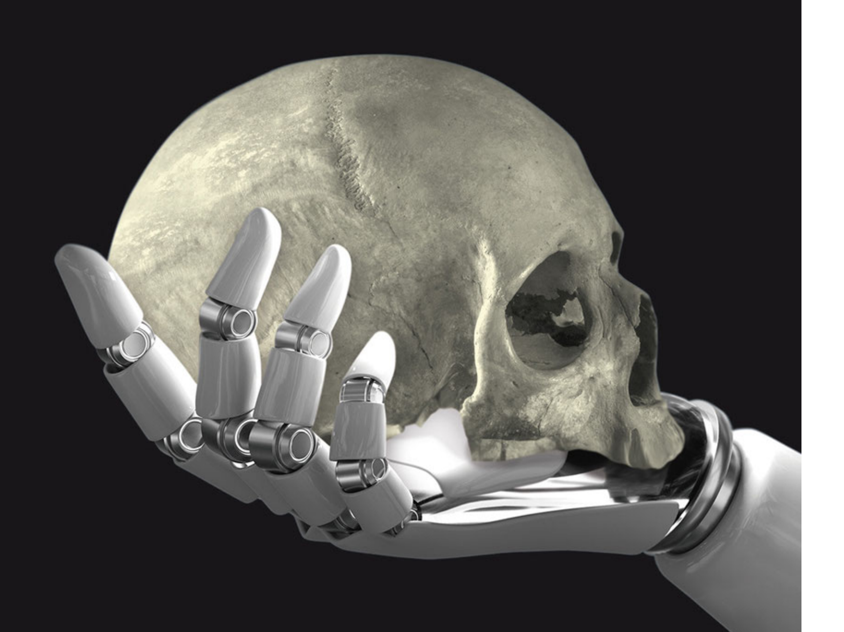




FERNANDO BONETE VIZCAÍNO

LA GUERRA IMAGINARIA

Desmontando el mito de la inteligencia artificial con Asimov





La guerra imaginaria

Desmontando el mito de la inteligencia artificial con Asimov

La guerra imaginaria

Desmontando el mito de la inteligencia artificial con Asimov

Fernando Bonete Vizcaíno

Índice de contenido



españa
siglo xxi editores
www.sigloxxieditores.com
clave intelectual
www.claveintelectual.com
calle recaredo 3, 28002, madrid

argentina
siglo xxi editores
www.sigloxxieditores.com.ar
clave intelectual
www.claveintelectual.com.ar
guatemala 4824, c1425bup, buenos aires

méxico
siglo xxi editores
www.sigloxxieditores.com.mx
cerro del agua 248, romero de terreros, 04310, ciudad de méxico

Calle Recaredo, 3 - 28002 Madrid

editorial@sigloxxieditores.com

www.sigloxxieditores.com

Diseño interior: Sebastián Sánchez Yáñez

Diseño de cubierta: Sebastián Sánchez Yáñez

Corrección: Lola Delgado

Primera edición en España: mayo de 2024

ISBN: 978-84-323-2105-4

Digitalización: Proyecto 451

Para Maitane, siempre

Para Ignacio y Santiago

—Discúlpeme. ¿He comprendido bien su nombre? ¿Doctora Susan Calvin?

–Sí, señor Byerley.

—Es usted la psicóloga de «U. S. Robots», ¿verdad?

—Robopsicóloga, por favor.

—Oh, ¿acaso los robots, mentalmente, son tan diferentes de los hombres?

—Como la noche y el día —dijo ella, y se permitió una sonrisa glacial—. Los robots son esencialmente decentes.

«La prueba», en Visiones de robot

He observado que los hombres capacitados siguen estando solicitados en nuestra sociedad; todavía necesitamos al hombre que es suficientemente inteligente como para pensar en las preguntas adecuadas.

«Se puede evitar el conflicto», en Visiones de robot

Isaac Asimov

Introducción

L

os humanistas del apocalipsis tecnológico

Estamos en guerra con la inteligencia artificial y los robots. El cine y las series han hecho mucho daño con sus oscuras visiones de un futuro donde los robots son capaces de formular las reflexiones existenciales más complejas y las más hondas preguntas por el origen mismo de su existencia, antes de aplastar a una humanidad embrutecida (Ex Machina; Westworld); con su elucubración acerca de la existencia de proyectos informáticos secretos donde la computación cuántica es capaz de acongojarnos viendo la crucifixión del mismo Dios en directo y alterar la sucesión de acontecimientos que llamamos historia, para formular un nuevo presente a la carta (Devs); con su aterrador bestiario de mascotas dotadas de una inteligencia perversa («Metalhead», Black Mirror) y de los más violentos instintos asesinos («La ventaja de Sonnie», Love, Death & Robots); y con la sobrecogedora alteración de nuestras mentes tras desagradables intervenciones quirúrgicas o protésicas que extirpan una parte de nuestra vida (Severance; Peripheral).

Una mayoría de ensayistas también ha hecho mucho daño. Amparándose en el imaginario colectivo procedente de la ficción cinematográfica, seriéfila o literaria, han escrito lo que han querido –lo que han querido– pasar por una «no ficción» rigurosa sobre el catastrófico impacto futuro de la inteligencia artificial y la robótica. Respaldados por el despliegue de una aproximación humanística o una actitud filosófica ante el tema, sus obras acaban por generar imágenes tanto o más creativas y fantasiosas –en el sentido de enormemente exageradas– que la ficción.

Estos escritores han hecho del recelo frente a la inteligencia artificial y la robótica, sostenido por la industria audiovisual y el mesianismo de la industria tecnológica, un tema sobre el que fundar las peores predicciones del futuro, así como una fabulosa herramienta para el impulso de sus artes adivinatorias. Afirman que nos encontramos a las puertas de un apocalipsis digital, que nuestra civilización está al borde de su extinción, que el control del algoritmo ha creado un estado policial de mayor calado y gravedad que los grandes totalitarismos del siglo XX, o que la revolución tecnológica en marcha en nuestros días es la mayor que ha vivido la humanidad. Para ello no tienen reparo en adulterar el objeto de su crítica sumando a la discusión elementos más o menos ajenos a la inteligencia artificial y la robótica, como la cristalización del capitalismo más salvaje, el advenimiento de los movimientos de extrema derecha, o la crisis de la democracia liberal.

No se puede decir que no pueda tener lugar alguno de los supuestos anunciados, pero tampoco se puede dar por seguro que vayan a ocurrir, y menos todavía en un ejercicio de predicciones con apariencia de infalibilidad, pero sostenido sobre impresiones personales, que no sobre relaciones de causa o correlación. Se trata de lo que Erik J. Larson, científico experto en computación –ha trabajado para proyectos punteros financiados por DARPA– ha denominado «el mito de la inteligencia artificial», que «consiste en afirmar que su llegada es inevitable, mera cuestión de tiempo –que nos hemos adentrado ya en el sendero que conducirá a una IA de nivel humano, y más tarde a una superinteligencia–. No es así. Ese sendero existe solo en nuestra imaginación» (1). La magnitud y el vértigo provocado por esos futuribles e inferencias sin mediación de evidencias mínimamente rigurosas implantan la semilla de un miedo cerval en la ciudadanía y ofrecen argumentos sin mediación de pruebas para estímulo de los ilusionistas, negacionistas y conspiranoicos tecnológicos.

Desde luego, lo que seguro no se puede decir, por no corresponderse con la realidad, es que la humanidad ya ha quedado obsoleta o está a punto de ser desbancada por las máquinas. Poca credibilidad puede recibir quien afirma esto con absoluta seguridad y toda rotundidad en un momento en el que los coches autónomos apenas pueden aparcar en condiciones –ya no digamos salir a la

carretera sin causar un caos fúnebre–; en el que la versión de pago de ChatGPT no es capaz de ofrecer lo que se le pide cuando se sube un ápice el nivel de las instrucciones –no capta el contexto ni el propósito final por mucho que se le insista– y además limita su uso por usuario a 25.000 palabras, tras las cuales hay que sentarse a esperar más de tres horas para volver a utilizarlo; en el que la mayor parte de las herramientas de inteligencia artificial generativa de imagen y vídeo a nivel usuario –que no las producciones asistidas por genios informáticos que sacan en las noticias– presentan una calidad paupérrima, además de una capacidad creativa muy limitada; en el que los recursos educativos guiados por la inteligencia artificial consumen más tiempo en la preparación de cualquier material docente medianamente digno que el que lleva crearlos con medios informáticos convencionales; en el que no existe ningún robot humanoide que sostenga su farsa más de dos segundos.

Un momento en el que, en definitiva, no existe todavía ninguna inteligencia artificial que sea verdaderamente inteligente. Existen, claro que sí, inteligencias «estrechas» –Narrow Artificial Intelligences– que realizan tareas específicas con perfección inaudita e incluso muy superior al ser humano. Los casos son célebres en complejos juegos como el ajedrez –donde Deep Blue fue capaz de vencer al campeón del mundo Garri Kaspárov– o el go –donde AlphaGo (ahora AlphaZero) batió 4 a 1 al jugador surcoreano de 9° dan Lee Sedol–. Sin embargo, estas supercomputadoras son buenas haciendo la tarea específica para la que han sido diseñadas, y ninguna otra. Son súper máquinas, pero también súper limitadas, y su modelo algorítmico no puede escalarse para obtener del mismo una inteligencia general. Es decir, no existe ni estamos remotamente cerca de dar con una inteligencia artificial «general», aquella capaz de resolver problemas aleatorios en cualquier ámbito de acción, dotada de sentido común y de una capacidad eficiente y totalmente autónoma para aprender, razonar y planificar. (2)

Dado que un robot similar al que suele aparecerse en nuestra imaginación –un robot digno de la gran pantalla– debería integrar una inteligencia artificial fuerte para poder desempeñarse con una cierta habilidad y solvencia, pero al mismo tiempo esta inteligencia artificial no es factible en estos momentos, tampoco estamos en el

punto de poder disponer de un robot completo, capaz de manejarse con tareas de distinta índole y adaptarse a entornos cambiantes.

Y aun con todo, ¿por qué iban las máquinas a querer reemplazarnos? Como ha puesto sobre la mesa en reiteradas ocasiones la profesora e investigadora en Ciencia Cognitiva del departamento de Informática de la Universidad de Sussex, Margaret A. Boden, una de las más reputadas especialistas en inteligencia artificial del mundo, ni los robots ni la inteligencia artificial tienen objetivos, deseos o ambiciones. (3)

El esfuerzo de los humanistas del apocalipsis tecnológico por plantear la extinción de la humanidad como consecuencia ineludible de la aceptación y uso de la inteligencia artificial y la robótica solo es equiparable, en su imposible carrera por predecir el futuro, al ardor de quienes, desde el otro lado claman que el nuevo edén tecnológico para una nueva raza de seres pos o transhumanos «está cerca» (4). En los extremos de la irrealidad se encuentran los adversarios. Ambas predicciones resultan profundamente antihumanistas.

Por fortuna, la literatura de ficción especulativa contemporánea ha sido más ponderada y protagoniza, entre las obras artísticas destinadas a alimentar el contexto cultural de nuestros días, la excepción ante los excesos de negatividad y de positividad de una y otra corriente. Sin dejar de proyectar las pesadillas que nos provocan estas máquinas, sus páginas también hacen sitio a los sueños y anhelos de una sociedad que quiere ser mejor y ve en la inteligencia artificial y la robótica una herramienta valiosa para lograrlo, intentarlo o, al menos, ve en la confrontación con la máquina la posibilidad de reflexionar sobre lo humano para hacernos más humanos en un mundo que quiere serlo cada vez menos. Y no por culpa de las máquinas, sino de nosotros mismos.

Este planteamiento lo han liderado desde el relato un nutrido grupo de escritores chinos, en lo que se ha conocido como la nueva ola de la ciencia ficción china –Ken Liu (El zoo de papel, La chica oculta) y Cixin Liu (Sostener el cielo) entre los más conocidos–. También algunos pocos occidentales, entre los mejores, todo un Premio Nobel como el británico Kazuo Ishiguro (Klara y el sol), el estadounidense Ted Chiang (Exhalación) –ambos con significativas

raíces orientales–, el neerlandés de raíces chilenas Benjamín Labatut (Maniac) y, en menor medida –o con peor fortuna literaria– el también británico Ian McEwan (Máquinas como yo).

El equilibrio entre los miedos y las esperanzas que plantean todas estas historias alrededor de la robótica y la inteligencia artificial siempre ha definido a la mejor ciencia ficción, pero también al mejor pensamiento crítico. Que esta actitud se encuentre ubicada en la literatura, y no en la espectacularización de la taquilla o en los cenáculos de los intelectuales más tristones tiene su razón y origen en una de sus formulaciones más completas, la del humanista y escritor estadounidense Isaac Asimov.

Asimov escribió su primer relato de robots en 1939, con diecinueve años, y no dejó de escribir sobre ellos nunca, durante sus más de cincuenta años de carrera literaria. Él mismo acuñó el término «robótica» por primera vez en la historia, convirtiéndose así en el padre fundador de la misma. A través de esta «ciencia que estudia a los robots», que el escritor quiso desplegar con la ficción, Asimov ofrece respuestas a todas las grandes preguntas sobre el tema.

Lo hace, sobre todo, con una de sus series literarias más conocidas, el denominado «ciclo de los robots»: conjunto de relatos, novelas y muy contados pequeños ensayos en los que Asimov vuelca sus ideas acerca de la naturaleza de los robots y de la inteligencia artificial, los principios que guían su comportamiento y las claves sobre la pugna y convivencia entre máquinas y humanos.

La posición de Asimov sobre el impacto de las máquinas supuestamente dotadas de inteligencia –artefactos que, como se verá más adelante ya formaban parte de la realidad científica de su época–, hizo frente al planteamiento pro apocalipsis que, desde el momento fundacional mismo de la aparición de los robots en el imaginario colectivo moderno, desde la literatura en la década de los veinte del siglo pasado –y aun en sus prefiguraciones literarias decimonónicas– había fundado el modelo interpretativo sobre el rol destructor y solo aparente o engañosamente benigno que venían a desempeñar los robots en el devenir de la humanidad.

La interpretación generalizada acerca de los robots no siempre fue estrictamente negativa. El primer registro de un robot en la historia

de la humanidad lo encontramos en la mitología griega: Talos, el androide creado por Hefesto con la misión de proteger la isla de Creta.

Aunque este autómata de metal con forma humana es conocido en la historia de la literatura y del pensamiento por su aparición en los momentos finales de Las Argonáuticas de Apolonio de Rodas –o, para los amantes del cine clásico, por la animación en stop motion de Ray Harryhausen en Jasón y los argonautas (1963)–, tuvo su primera aparición conservada en Simónides, en fechas tan tempranas como el siglo VI y V a. C. Allí se señala a Talos como un phylax empsychos, un «guardián animado», sobrenombre con el que se designa su tarea protectora y, por tanto, bondadosa. (5)

Sin embargo, para Medea y los argonautas que sustrajeron el vellocino de oro de la isla, este guardián es un enemigo declarado, dado que, en cumplimiento de la misión que le otorgó su creador, intenta acabar con ellos lanzándoles piedras. En esta primera comparecencia del robot en la mentalidad humana, se manifiestan sus virtudes, tanto como sus peligros.

Peligros que Asimov reconoció en el momento de evaluar los derroteros por los cuales podían conducirnos la robótica y la inteligencia artificial. Nunca esquivó esos debates en sus ficciones. Muy al contrario, su narrativa está repleta de situaciones y entuertos en los que resolver debates éticos propiciados por la evolución tecnológica –sucesos que ahora nos servirán de apoyo en este libro para sostener la discusión sobre el tema–. Pero planteó estos escenarios, sabedor de que todos y cada uno de los mecanismos creados o utilizados por el hombre tienen sus peligros y son susceptibles de ocasionar problemas, tanto como de protegernos y ayudarnos a progresar.

Con ejemplos más o menos provocadores, Asimov nos reta a pensar en esos otros mecanismos que pueden aportar tanto como entorpecer en la tarea de hacer progresar la humanidad:

El descubrimiento del lenguaje introdujo comunicación –y mentiras–. El descubrimiento del fuego introdujo la cocina –y el

incendio–. El descubrimiento de la brújula mejoró la navegación –y destruyó civilizaciones en México y Perú–. El automóvil es maravillosamente útil –y mata decenas de miles de norteamericanos cada año–. Los adelantos médicos han salvado millones de vidas –e intensificado la explosión demográfica–.

En cada caso, se pueden utilizar los peligros y abusos para demostrar que «hay ciertas cosas que la Humanidad no estaba destinada a conocer», pero sin duda no se puede esperar que renunciemos a todos los conocimientos y volvamos al estado del australopiteco. (VR, 19-20)

Con afán de compensar el pesimismo irracional imperante en el canon literario de su tiempo –y en la búsqueda de una originalidad que todavía hoy no ha perdido–, Asimov decide que en su literatura no impere la imagen del robot malvado, o que, al menos, no haya robots con autonomía para obrar el mal; ni mucho menos superinteligencias con planes maléficos capaces de coordinar a todas las máquinas en contra de la humanidad. En todo caso, encontramos personas malas que intentan utilizar para su propio beneficio y otros fines inmorales a los robots. Pero siempre menos que personas buenas –acompañadas de buenos robots– dispuestos a desenmascarar la farsa y devolver la robótica y la inteligencia artificial, cuando no la sociedad en su conjunto, a la senda del bien y del servicio a la humanidad.

En buena medida, la tesis defendida por Asimov fue el primer aldabonazo contra los detractores de la robótica y la inteligencia artificial y el pesimismo de sus predicciones tecnológicas, que pretendían conocer «el día y la hora» de la parusía robótica. Desde la década de 1940, este fin de los tiempos desatado por las máquinas iba a ocurrir, supuestamente, a dos décadas vista. Siempre va a ocurrir a dos décadas vista. Sin embargo, pasadas ocho de esas décadas, el vaticinado y esperado fin de los tiempos propiciado por la conquista de los robots o la dictadura globalista de una superinteligencia computerizada no ha llegado, y la evidencia científica de que algo así pueda tener lugar en el corto o medio plazo es más bien inexistente –porque en el largo plazo, ese terreno de los oráculos y las artes adivinatorias, todo y nada puede

pasar–.

Pese a todo, la intensidad de las predicciones más aciagas se ha redoblado. De hecho, se ha acortado en términos temporales, pues muchos autores claman que el punto de no retorno ha llegado y el ser humano ha quedado obsoleto. Es evidente que el hecho de que con anterioridad esos vaticinios fueran erróneos no significa que ahora también lo sean, pero sí significa que podrían volver a serlo. La historia sienta precedente.

Con todo, el alegato de Asimov, formulado con toda la fuerza de los clásicos y de un genio literario bien nutrido de conocimientos científicos, históricos y filosóficos –no solo filosóficos, no solo históricos, no solo científicos–, alegato en defensa de un juicio sobre el tema que no obvie el verdadero estado del desarrollo técnico de estas tecnologías y no elimine de la ecuación las enormes potencialidades del ser humano sigue vigente y merece ser recuperado para arrojar luz a los debates del presente.

Cuando un lector se enfrenta a un clásico de la ciencia ficción, corre el riesgo de deslumbrarse por lo que, en su natural veneración hacia el autor y exceso de ciega fascinación, percibe como extrema «clarividencia» del mismo para «profetizar» o «vaticinar» los adelantos científicos y tecnológicos que solo décadas más tarde serán una realidad. La lectura de Asimov no está exenta de este riesgo.

El escritor nos habla de cuatro (como tarde) a ocho (como pronto) décadas vista de realidades tecnológicas como la inteligencia artificial, la robótica o la cibernética, realidades que, por entonces, en el momento de escribir y publicar su obra no gozaban del desarrollo que presentan hoy y, lo que es más importante para la integración de estas realidades en el imaginario del lector, tampoco formaban parte de la experiencia diaria de un público no especializado.

Hoy todo el mundo puede hacer uso –o hace uso sin saberlo– de la inteligencia artificial; la implementación de maquinaria robótica forma parte de la cadena de valor de la industria en la práctica totalidad de los sectores; y es cada vez más raro no disponer de un robot en casa –siquiera en su forma y funciones más simples–. Por

entonces, cuando Asimov escribía su obra de ficción en relación a estos o similares sistemas y artefactos, ninguno de ellos resultaba ni remotamente familiar para una vasta mayoría de las personas. Pero existían.

Existían y, como veremos a continuación, se hicieron grandes avances teóricos y prácticos en los momentos en los que Asimov escribió su obra de ficción al respecto. Es decir, Asimov no partió ex novo en sus planteamientos y consideraciones sobre estos temas. Fue el primero en acuñar el término «robótica», pero la naturaleza del término aduce a realidades que ya tenían un desarrollo previo: los trabajos sobre los «sistemas autoorganizados», realizados en torno a las mismas fechas en las que él escribe sus relatos sobre robots o, desde una perspectiva histórica, los autómatas humanoides atribuidos a Leonardo da Vinci a finales del siglo XV, los artilugios de relojería de la Edad Media o, con anterioridad, los artefactos automotrices del período helenístico –por no decir de la presencia de «vida artificial» en los mitos griegos–.

Es cierto que el denominado «cerebro positrónico» con el que funcionan los robots de sus ficciones, y las extraordinarias capacidades que poseen sus máquinas gracias a este artefacto, no han sido alcanzadas, ni de lejos, siquiera por las complejas redes neuronales artificiales de nuestros días, todas ellas de una simpleza apabullante si las comparamos con el cerebro artificial proyectado por Asimov, y son muy pobres frente a las potencialidades del cerebro humano, pero aun con todo, la incorporación de la inteligencia artificial a los sistemas autoorganizados data de la década de 1950.

Deteniéndonos un tanto en este asunto, sí es cierto que en esto del cerebro positrónico Asimov se adelantó tanto al futuro que ni siquiera ahora, en lo que llevamos de siglo XXI y en el momento de escribir estas líneas existe algo que iguale o se asemeje remotamente a la complejidad y funcionalidades de ese artefacto de ficción. Tampoco podemos saber si existirá en el futuro, pese a que haya quienes lo den por sentado, en un ejercicio de la imaginación que solo puede tener su correlato en el cine, las series o la literatura, es decir, en la ficción. Como mucho, de lograrse una proeza tal, tendrá que ser en un futuro muy lejano, pues la

evidencia disponible hasta el momento nos muestra que los proyectos con financiación billonaria en inteligencia artificial y neurociencia desplegados en países de todo el mundo con el objetivo de emular un cerebro completo –Whole Brain Emulation– ofrecen resultados escasos y a cuentagotas. Esta circunstancia no tiene por qué resultar extraña pese a la cantidad de años transcurridos y de recursos invertidos si se considera que estos proyectos tratan de replicar un cerebro humano, objetivo en el fondo sorprendente teniendo en cuenta que ni siquiera conocemos el cerebro humano en su totalidad.

Gracias a las avanzadas capacidades de un cerebro positrónico, el grado de perfección de la inteligencia artificial permitiría a la práctica mayoría de los robots avanzados de Asimov –los no dedicados a tareas rutinarias de trabajo básico– pasar un test de Turing sin problemas, algo no imposible, pero mucho menos probable con el estado actual de desarrollo tecnológico en la materia.

Además, los robots de Asimov también suponen una novedad con respecto a su soporte físico, dado que en nuestros días –a pesar de lo que dé a entender el marketing y la publicidad– la gran mayoría de las generaciones de robots en desarrollo no son físicos, y evolucionan en modos de simulación, sin duda para rentabilizar costes. A algunos de esos robots, pocos, pero los más icónicos de su narrativa, Asimov los reviste de aspecto humano externo; son humaniformes, indistinguibles de una persona para el ojo humano sin una observación exhaustiva, y también pasarían un test de Turing visual si tal cosa existiera.

Volviendo a la cuestión del estado del arte en los avances tecnológicos de su tiempo y cuánto se adelantó Asimov en realidad a nuestros días, para cuando escribe buena parte de sus relatos de robots, inclusive su novela Bóvedas de acero, el escritor integra y hace notar con claridad la presencia y reflexiones acerca de la inteligencia artificial en un momento –previo a 1956– en el que el célebre informático John McCarthy todavía no había acuñado el término. Pero se la conocía como «simulación computerizada»; había sido vaticinada por Ada Lovelace en la década de 1840; había sido demostrada y trabajada por Alan Turing en 1936, refrendada

por las investigaciones del neurólogo y psiquiatra Warren McCulloch y el matemático Walter Pitts en los cuarenta y, para el momento en que escribe Asimov, recibía su primer impulso tecnológico fuerte con máquinas más potentes y nuevos lenguajes de programación que harían posible la Máquina de la Teoría Lógica –demostró dieciocho de los teoremas lógicos principales de Bertrand Russell– o el todavía más potente Solucionador General de Problemas, capaz de solucionar cualquier problema representado en términos de objetivos y operaciones. Para finales de 1960 ya se estaban manejando todos los tipos principales de inteligencia artificial que hoy conocemos.

No cabe duda de que Asimov conocía los avances realizados en materia de robótica e inteligencia artificial. Quizá no desde un punto de vista técnico o especializado, puesto que, de hecho, rechazó la propuesta del sello Harmony Books de escribir en solitario Robots. Máquinas a imagen y semejanza del hombre, un ensayo sobre la historia, desarrollo y aplicaciones de los robots (6), pero sí conocía y disponía de las suficientes nociones para utilizarlos como material de sus obras de ficción con inteligencia e interés y plantear los temas y preguntas clave. Aunque la inspiración literaria sigue complejos derroteros, rara vez se sustrae al espíritu del tiempo al que pertenece.

Con todo, la mayor virtud de Asimov en el momento de valorar el componente tecnológico en su obra no está tanto en la novedad material y funcional de los artefactos que comparecen en su narrativa, pues en su mayoría y a rasgos más o menos generales gozaban de plena vigencia en su tiempo. Su mayor virtud reside en el vigor de los comentarios y reflexiones acerca de esos artefactos. Asimov nos habla desde el pasado con toda actualidad, porque su obra contiene el mensaje universal que sirve para identificar a los grandes clásicos.

En este libro se han planteado tres grandes bloques de discusión sobre inteligencia artificial y robótica siguiendo criterios de actualidad e interés con respecto a su manifestación en el debate público. En todos ellos comparece el pensamiento de Isaac Asimov como la base del tratamiento y guía para la reflexión de cada una de las cuestiones tratadas en estos tres grandes capítulos.

El primer bloque concierne a nuestras expectativas acerca de la relación que mantenemos o deberíamos mantener con los robots y la inteligencia artificial, e intenta responder a las preguntas más habituales sobre el tema: ¿de qué forma hemos de tratarlos? ¿De qué forma han de tratarnos ellos a nosotros? ¿Cómo se articula en la teoría y en la práctica la relación entre el creador y su creación? ¿Pueden llegar a suponer una amenaza real para la humanidad? ¿Cuál es el origen y motivación del miedo que albergamos frente a estos artefactos? ¿Acaso este miedo alberga algún correlato con la realidad de los avances técnicos de estas máquinas y sistemas de computación o es, por el contrario, infundado? ¿Debería existir un sistema ético universal que regulara el comportamiento de estas máquinas? ¿Qué sistema ofrecería suficientes garantías de control?

El título de este capítulo, «El buen robot», quiere ser toda una declaración de intenciones, negando la máxima que lleva a considerar a los robots, junto a la inteligencia artificial, la gran amenaza para la supervivencia de la humanidad. Una reacción y un sentimiento naturales, si bien en el fondo irracionales e injustificados a los que el propio Asimov dio el nombre de «complejo de Frankenstein».

Al contrario de lo que nos hace creer el marketing de las empresas tecnológicas detrás de estos servicios –que necesitan cuantas más suscripciones a su app mejor–; al contrario de lo que nos hace creer la incesante aparición de supuestas proezas robóticas en los medios de comunicación –que necesitan sorprendernos a toda costa para alcanzar el ratio de clics y visitas esperado, sin importar si para ello tienen que trucar vídeos e incorporar efectos especiales para simular que los robots no solo disparan, sino que pueden hacerlo mientras ejecutan acrobacias–; al contrario de lo que nos hacen creer los anuncios de revolución –todos los días tendría lugar una revolución distinta– de los gurús tecnológicos –que necesitan mantener el flujo de views y likes–; pero también al contrario y en contra de las peores y más exageradas predicciones proferidas por los humanistas del apocalipsis tecnológico sobre el insalvable y desastroso impacto que están ocasionando y van a ocasionar estas tecnologías. Al contrario, y en contra de todas estas exageraciones, las posibilidades que ofrecen la informática y la fuerza de computación disponible en estos momentos son, en realidad,

demasiado limitadas para alcanzar el ideal de inteligencia y completa autonomía falsamente atribuida a estos sistemas; insuficientes para causar la tan deseada debacle y llevarlos a hacerse con el control del mundo.

Hay en los logros de la inteligencia artificial y la robótica un alcance y una magnitud mucho más reducidos de lo que se nos hace creer para vender, y unos fallos más que recurrentes en la ejecución de las tareas que los humanos les encomendamos nunca mencionados. Unos con objeto de sostener una interesada apariencia de progreso, y otros con la candorosa aspiración de interrumpir el curso de la historia y mantener inalterado el estado de las cosas para seguir teniendo algo que decir sobre ellas. En principio, los esfuerzos por criticar siempre son más rentables que el interés por conocer.

Por su parte, el segundo y el tercer bloque de esta obra presentan aquellas habilidades y competencias propias de las máquinas: qué características las definen y de qué son capaces. La orientación y propósito de esta indagación es, sin embargo, distinta para cada uno de estos apartados.

En el segundo, tal y como indica su título, se quieren desentrañar las propiedades y características de la máquina y el posible alcance de sus logros para, de su reflejo, pero sobre todo de sus ausencias y carencias, es decir, de lo que no es y no puede ofrecer, apuntar a una respuesta para la pregunta «¿Qué es lo humano?»: qué es lo nuestro y de ningún otro.

Se procede así, por sustracción, a determinar las potencialidades exclusivas de las personas, señalar aquello reservado al ser humano, que no puede ser siquiera imitado por la máquina y es constitutivo de nuestra manera de desenvolvernos y de proceder en la vida y que nos hace, en definitiva, insustituibles.

En buena medida, esta identificación de las diferencias sustanciales entre el hombre y la máquina procede de las relaciones que establecen dos de los más icónicos protagonistas de la ficción de Asimov, el detective Elijah Baley y la robopsicóloga Susan Calvin, con los robots que habitan las páginas de su narrativa. Ambos personajes se dedican a resolver problemas, el primero de índole

criminal y la segunda de índole psicológico, con lo que sus ocupaciones están perfectamente delineadas para generar el tipo de casos, problemas y dilemas capaces, no solo de mantener el interés de la trama, sino de plantear las preguntas desde las que estimular al lector –al lector de estas novelas, al mismo tiempo que al lector de este libro– para involucrarse en el tema.

En el tercer capítulo, «Los riesgos del progreso robótico», esas mismas particularidades de la máquina y sus diferencias o limitaciones frente a la naturaleza humana se utilizan con la pretensión de reflexionar sobre los riesgos que conlleva, bien su desconocimiento, bien una comprensión errónea, un mal uso o una aplicación de las mismas como falsas sustitutas de las capacidades de las personas. En este punto, las proyecciones de distintos futuros posibles llevadas a cabo por Asimov en sus novelas, en forma de mundos espaciales en los que sus habitantes han sido conquistados por una mentalidad robótica, son sumamente ilustrativas de la deshumanización que puede ocasionar el olvido de lo que nos hace humanos.

Con todo, en este libro no se obvian los retos que plantean la inteligencia artificial y la robótica en un momento de grandes cambios y creciente presencia de la tecnología en nuestras vidas. Tampoco se elude el riesgo que un sobredimensionamiento y descontrol de ambos puede tener en la configuración misma de nuestra vida. Sin embargo, en estas líneas no hay sitio para el paroxismo. Ni la humanidad ha quedado obsoleta ni está condenada a desaparecer o a extinguirse por culpa de los robots. O no tiene por qué estarlo.

Acompañados de la visión humanística del padre de la robótica, planteamos una revisión de los temas clave de la inteligencia artificial contando con que sus peligros existen y sus consecuencias negativas son una posibilidad, pero nunca un destino manifiesto e inexorable.

▪

1 Larson, El mito de la inteligencia artificial, p. 7.

2 Por utilizar la definición que ha hecho de la misma precisamente uno de los mayores exponentes del mito de la inteligencia artificial, Bostrom, Superinteligencia. Caminos, peligros, estrategias.

3 Boden, Inteligencia artificial.

4 Kurzweil, La singularidad está cerca.

5 Mayor, Dioses y robots. Mitos, máquinas y sueños tecnológicos en la Antigüedad, pp. 1-34.

6 Con respecto al ensayo, la experta en el tema Karen A. Frenkel escribió la mayor parte del mismo en coautoría con Asimov –aunque este firmó como primer autor para favorecer el éxito de unas ventas que después no fueron tales– y el libro se acabó publicando en 1985.

El buen robot

Los robots son las máquinas que más se parecen al ser humano. No solo o no tanto por su apariencia visual; en ocasiones, la apariencia humanoide de un robot es incluso contraproducente, y provoca que percibamos antes sus diferencias que sus similitudes. Los robots son las máquinas que más se parecen al ser humano porque pueden realizar «tareas demasiado complejas para cualquier mente viviente aparte de la del hombre, y de unas características que una máquina no-computerizada no es capaz de realizar» (VR, 13).

Con estas pocas líneas sobre las funcionalidades del robot, contenidas en uno de esos breves ensayos que Asimov escribió sobre robótica y que gustaba de incluir junto a sus relatos sobre el tema, para facilitar la comprensión del papel desempeñado por los robots en sus tramas, el escritor nos recuerda que estos son tan capaces de realizar tareas complejas a nivel operacional y físico que, en no pocas ocasiones, sus habilidades específicas se sitúan por encima del nivel de un ser humano, incluso de uno especialmente brillante o habilidoso.

Este tipo de robots, capaces de sobresalir en actividades muy concretas y acotadas son una realidad cada día más frecuente. Sin embargo, no son el tipo de robot que puebla nuestra imaginación y alimenta nuestros temores. El robot que nos atemoriza es el robot «completo», el más perfecto de su clase, aquel nacido de la más refinada ingeniería, aquel que podría hacerse pasar por un humano y no ser detectado como robot. Aquel que se nos presenta como real, pero no existe todavía, pues resulta inviable para nuestros limitados conocimientos en computación. Aquel que solo es posible a través de las infinitas posibilidades de la ciencia ficción.

El tipo de robot imaginario, el único capaz de presentarse ante nosotros dotado de todas las capacidades y apariencia humana, no es, sin embargo, y nunca sería un ser humano. Los robots son

«como» nosotros. Entonces, los robots no son enteramente como nosotros; no son nosotros. Este parecerse y no ser, y la aspiración constante a igualar o superar al ser humano allí donde más destaca –mucho más que las características de la máquina en sí mismas y los prodigios de los que pueda ser capaz– son la fuente principal de los recelos y el miedo que, por lo general, nos suscitan los robots.

Miedo a lo similar extraño; al ente máquina que se nos presenta con una apariencia familiar, a su vez repleta de carencias que nos hacen echar en falta su humanidad o percibirla en grado sumo –la falta o exceso de humanidad en un mecanismo que emula la humanidad–. Miedo a que un ingenio mecánico posea las facultades en principio reservadas a los seres humanos. Miedo a que la relación de nuestra especie con máquinas dotadas de estas habilidades, en grado similar o superior al nuestro degenere en una rivalidad en la que seamos borrados del mapa o sustituidos. Al respecto decía Asimov: «Seguramente el miedo mayor no es que la maquinaria nos lastime, sino que nos suplante» (VR, 505).

Sin tener que ir tan lejos, en realidad basta con el miedo a que nuestra inferioridad sea señalada o evidenciada para que se genere el rechazo. Así se refleja en los temores de Herbie (RB-34) (7), el robot protagonista del relato ¡Mentiroso!, cuando desvela la razón por la cual no puede contar la verdad a sus dueños y se ve abocado a mentirles: «Ustedes no pueden perder la cara delante de mí sin sentirse heridos» (VR, 133).

En esta ficción, el robot, afectado por un error de fábrica, crea una red de mentiras entre sus interlocutores humanos; con sus embustes proyecta o da continuidad a sus ilusiones para evitar su decepción y evitar que lo pasen mal. En definitiva, Herbie no puede contar la verdad, porque la verdad deja al descubierto nuestra vulnerabilidad frente a la realidad, y puede llegar a generar una frustración que se vuelve contra quienes ponen –o lo que pone, en este caso– dicha verdad sobre la mesa.

Ante esta situación es imposible no pensar en el tipo de respuestas que ofrecen herramientas de inteligencia artificial generativa de uso diario como ChatGPT, Copilot o Gemini. Cuando se les pregunta acerca de una cuestión ética de importancia, estas aplicaciones suelen optar por mantener una estudiada equidistancia. Esta

neutralidad acaba por hacer de su posición, en la práctica, una aportación insustancial, pero evita que nos sintamos atacados por una actitud contraria a nuestras creencias. Ideas contrarias a las nuestras que, para más inri, ni siquiera son sostenidas por otra persona, un igual, sino por una máquina cuyos propietarios, además, dependen de las suscripciones de unos usuarios –nosotros– a los que deben evitar molestar a toda costa para que sigan sosteniendo el negocio con sus aportaciones económicas.

En otras ocasiones, estos sistemas pueden optar directamente por la invención y la mentira. Preguntado por una realidad de índole teórica, ChatGPT elabora para el usuario una explicación detallada de la misma recurriendo al conocimiento obtenido durante su entrenamiento con millones de unidades de artículos científicos. Una vez cargada la respuesta, se le puede requerir un listado de referencias en las que se haya basado para elaborar toda o parte de la misma. En este escenario, no es infrecuente que la herramienta malinterprete el contexto y prime la orden del usuario frente al rigor académico que se le presupone –una presunción errónea– o frente a valores de autenticidad, originalidad y veracidad –valores que, en sí mismos, una máquina no posee–, e invente fuentes de información para respaldar la respuesta. El usuario puede encontrarse entonces sumando a sus informes o trabajos referencias incoherentes con el contenido del texto, falsas o directamente inexistentes.

Con el ejemplo de Herbie, ese defectuoso, pero elocuente robot imaginario de 1941, Asimov nos muestra a ochenta años vista lo poco que hace falta para que odiemos a un robot, si bien al mismo tiempo este ejemplo narrativo nos recuerda que ese poco que hace falta es el detonante universal de todas las grandes incomprensiones –la verdad hiere, la verdad molesta– y nos pone ante una diatriba sumamente interesante en las comparaciones hombre-máquina. El ser humano es el único ser capaz de atender con autenticidad y dar un sentido completo a esta cualidad. Si perdemos nuestra relación con la verdad, ¿no estamos más cerca de convertirnos en máquinas? Parece una cuestión que merecería la pena reflexionar en tiempos de la posverdad.

El creador se vuelve contra su creación

La actitud de sospecha hacia los robots, la voluntad de humanizarlos o la pesadilla de su posible gobierno sobre la humanidad son posiciones que han imperado desde la misma aparición de la palabra que les da nombre, en 1920.

No fue en el ensayo, sino en la ficción, en la obra de teatro R.U.R. del escritor checo Karel Čapek (8) –las siglas del título responden a Robots Universales Rossum, la empresa que los fabrica–, donde está el origen del término, la palabra checa «robota» que significa «esclavo» y resume bien el fondo de esas actitudes primerizas, luego tan replicadas y extendidas en la ficciones que estaban por venir: la creación de una fuerza de trabajo incansable, hiperexplotable, con grandes capacidades productivas y dotada de inteligencia que acaba por adquirir intereses propios y se rebela contra sus creadores para forjar un mañana gobernado por ella.

Muchos otros autores volvieron a partir de entonces sobre este papel amenazante desempeñado por los robots, una fórmula narrativa que sigue vigente hoy en ficciones con todo tipo de formatos y géneros –estos autores, se quejaba Asimov, no tenían conocimientos científicos, y por eso eran hostiles a la ciencia–. El propio Asimov desdeñó a los creadores que abundaron en una idea tan manida, poco original y, como veremos a continuación, poco realista: «Solo una conjura de un robot parece ser la cosa más a mano para un autor medio: el hombre mecánico que demuestra ser una amenaza, la criatura que se vuelve contra su creador, el robot que se convierte en una amenaza para la Humanidad» (VR, 465).

La visión de Asimov acerca de la robótica es mucho menos catastrofista y dramática, más esperanzadora, pero, sobre todo, más realista por ser –y esto la mantiene vigente en nuestros días– una aproximación científica al tema. Sí, los robots pueden albergar riesgos y provocar problemas, pero ¿están dotados de las capacidades que puedan hacer de ellos un peligro fatal o una

amenaza insalvable para la humanidad? ¿Son ellos en realidad los causantes de los contratiempos que se les atribuyen? ¿Dónde se halla el origen, la causa primera de dichos peligros, amenazas, riesgos y problemas? Para Asimov está claro: en la naturaleza imperfecta de sus creadores, hombres y mujeres capaces de lo mejor y de lo peor; nunca en los robots, cuya existencia y acciones no son sino resultado dependiente de las aspiraciones y órdenes de las personas que los crean o los utilizan.

Esta idea es de suma importancia para combatir «el mito de la inteligencia artificial» y el discurso de aquellos pensadores que, ya sea con derivaciones sumamente pesimistas o escatológicas –apocalípticos tecnológicos–, o extremadamente optimistas –singularistas–, pero siempre influenciados por el mito, nos dan a entender o crean la impresión falsaria de que los robots y la inteligencia artificial son una especie viva más, con una línea evolutiva desgajada de sus creadores humanos y dotada de la autonomía necesaria para progresar por sí misma.

Esto, por supuesto, es totalmente falso –lo era entonces, lo es ahora, y la evidencia científica apunta a que lo seguirá siendo por mucho tiempo–, y toda la obra de Asimov en relación a la robótica –además de las célebres novelas integrantes del ciclo de los robots, escribió relatos sobre el tema durante casi cincuenta años–, con todos sus casos policiales, legales e investigaciones científicas articuladores de sus tramas, se empeñan en negar la mayor demostrando que, en primer y último término, los robots y la inteligencia artificial que les dota de impulso son en sí mismos seguros y fiables.

Para Asimov es absurda la idea de que un robot se vuelva contra su creador. Una máquina no se vuelve contra su creador si está bien diseñada, es decir, si su creador, que es quien la programa, inserta las órdenes que ha de cumplir la máquina y, de quien esta depende en todos sus órdenes, no quiere. Más bien ocurre al revés: los creadores se vuelven contra sus creaciones robóticas tanto como los usuarios de las mismas descargan su ira contra ellas.

En el fondo buscamos soliviantar nuestra desilusión vital sobre entes que son el fiel reflejo de nuestros defectos. Los robots, resultado de una acción humana que desea imitar lo humano con la

inalcanzable búsqueda de perfección que nos caracteriza –un fracaso que nos mueve, tanto como nos paraliza– acaban por dar lo que pretendíamos de ellos, una imitación de lo humano y, por tanto, una imitación de nuestra natural imperfección. El resultado de caer en la cuenta, directa o indirectamente, de la imperfección de la que son prueba ineludible nuestras creaciones es la frustración y el consiguiente desahogo contra ellas.

Algo así sugiere Asimov en la novela Bóvedas de acero, relato que nos sitúa en el escenario de las Ciudades –con mayúscula–. Son enclaves de la Tierra que simulan a las grandes ciudades existentes en el presente, si bien en la ficción han alcanzado dimensiones monstruosas –por ejemplo, 5.000 km^2 para la ciudad de Nueva York, que en nuestro presente tiene una superficie de 784 km^2–. Las Ciudades son núcleos hiperpoblados y subterráneos, multiplicados en extensión vertical con docenas de niveles bajo tierra. Hay viviendas situadas en el límite con la superficie, el denominado «espacio abierto» o Nivel Superior, pero son poco deseadas por estar más expuestas a las civilizaciones espaciales –sociedades robóticas– y su alquiler corresponde a personas muy poco pudientes.

En estas Ciudades viven hacinadas grandes masas de población en un disimulado régimen de corte autoritario. Una sociedad frustrada por sus escasos logros y una crisis política y cultural latente sirven de trasfondo al argumento de Bóvedas de acero, primera novela del ciclo de los robots que transcurre, al contrario que las partes continuadoras de esta saga, en la Ciudad de Nueva York.

En este régimen de descontento, son los ciudadanos quienes culpan a los robots, cuya presencia va en aumento en trabajos poco –pero cada vez más– cualificados, de sus fracasos como colectivo. El malestar social se proyecta sobre los robots solo porque es fácil dirigir contra ellos la ira acumulada:

Los hombres que se encontraban enfrentados a la perspectiva de recibir solo un mínimo de miseria si perdían su cualificación, después de media vida de esfuerzos, no podían decidir con la cabeza fría que un robot individual no era el culpable. Al menos, a los robots individuales se les podía pegar. No se podía pegar a algo

llamado «política gubernamental» o a una consigna como «mayor producción mediante el trabajo robótico». (BA, 34)

Asimov nos coloca ante la idea de que el robot «culpable» es tan solo un chivo expiatorio más, una víctima propiciatoria resultado de las frustraciones humanas. Las personas, en su incapacidad para tomar el control de sus vidas y hacer frente a los problemas de su tiempo atacándolos de raíz, prefieren tomarla con la máquina. Se alivia de manera momentánea la tensión social, pero no se soluciona nada.

Esta idea se manifiesta con fuerza en nuestros días, casi como una advertencia. A tenor de la reflexión anterior, no parece tan casual que en una época de grandes transformaciones en todos los ámbitos –una época de cambios para un cambio de época–, donde los descontentos en muy diversos órdenes abundan, pero la falta de capacidad crítica generalizada dificulta articular nuestros juicios y orientar nuestras opiniones con la profundidad y el tino debidos, lo fácil sea apuntar a estas creaciones tecnológicas disruptivas como la fuente de todos los desajustes presentes y de las amenazas futuras.

Esto aun cuando el nivel de desarrollo real de la robótica y de la inteligencia artificial es todavía muy limitado y las previsiones de dicho desarrollo para el mañana son más limitadas todavía, y aun cuando al final de la cadena de causas y correlaciones, el responsable de estas máquinas, así como del escenario del descontento, no es otro que el ser humano, es decir, nosotros mismos.

Por tanto, quienes no son seguros ni fiables son los creadores por el mal diseño que hacen de los robots, y las personas dueñas de esos robots por el uso irresponsable que hacen de ellos. La cara positiva de esta línea de pensamiento es que, si los robots están capacitados para hacer el bien, y en no pocas ocasiones se erigen en representación de la bondad frente a las personas, es porque el ser humano ha inculcado en ellos lo mejor de sí mismo.

Las Tres Leyes de la robótica

Las Tres célebres Leyes de la robótica, que Asimov consideró su más importante contribución a la ciencia ficción, fueron formuladas por primera vez en el relato El círculo vicioso (1941). Son quizá la expresión y el reflejo más favorable y positivo de las aspiraciones humanas insertadas en los robots.

De hecho, del primero de estos postulados y en representación de todos ellos, dirá Susan Calvin, la célebre robopsicóloga y heroína de Asimov: «A través de la contundente fuerza de la Primera Ley de la Robótica, [los robots] tienen en su interior la bondad de la Humanidad» (VR, 227).

1. Un robot no puede hacer daño a un ser humano, o, por medio de la inacción, permitir que un ser humano sea lesionado.

2. Un robot debe obedecer las órdenes recibidas por los seres humanos, excepto si estas órdenes entrasen en conflicto con la Primera Ley.

3. Un robot debe proteger su propia existencia en la medida en que esta protección no sea incompatible con la Primera y la Segunda Ley.

La misma robopsicóloga señala la evidencia que proporciona un estudio detenido de cada una de estas leyes: que las tres se corresponden con «los principios esenciales de una gran cantidad de

sistemas éticos del mundo» (YR, 308):

Todo ser humano se supone dotado de un instinto de conservación. Es la Tercera Ley de la Robótica. Todo ser humano bueno, con conciencia social y sentido de la responsabilidad, deberá someterse a la autoridad constituida; obedecer a su doctor, a su Gobierno, a su psiquiatra, a su compañero; aunque sean un obstáculo a su comodidad y seguridad. Es la Segunda Ley de la Robótica. Todo ser humano bueno, debe, además, amar a su prójimo como a sí mismo, arriesgar su vida para salvar a los demás. Esta es la Primera Ley de la Robótica. Para exponerlo claramente, si Byerley observa todas estas reglas, puede ser un robot, pero puede también ser simplemente una buena persona. (YR, 308-309)

El escritor Rodolfo Martínez (2021), estudioso de la narrativa de ciencia ficción de Asimov, ha equiparado con acierto las Tres Leyes de la robótica a las «leyes de comportamiento del buen ser humano», reformulándolas del siguiente modo (9):

1. Un ser humano no hará daño a otro ni permitirá, por inacción, que este sufra daño.

2. Un ser humano obedecerá las leyes vigentes, excepto si estas entran en conflicto con la Primera Ley.

3. Un ser humano preservará su propia existencia, excepto si esto entra en conflicto con la Primera o Segunda Ley.

Para Asimov, las Tres Leyes de la robótica vuelcan en los robots lo mejor del ser humano, y hacen que las máquinas ofrezcan una

protección perfecta para las personas, subrayando que los robots son totalmente seguros para la humanidad; estas leyes velan por la vida y el bienestar de las personas, y garantizan que estas se mantengan por encima de los robots en orden de importancia.

En el relato ¿Qué es el hombre?, Keith Harriman, director de investigaciones de U. S. Robots & Mechanical Men –la empresa ficticia que fabrica los robots en buena parte de la narrativa de Asimov– es preguntado por el robot George Diez (JG-10) si él siente el temor de ser reemplazado por los robots, temor del que adolece buena parte de la opinión pública (CC2, 620):

—Claro que no. Mientras existan las tres leyes de la robótica, es imposible. Los robots pueden actuar como socios de la humanidad; pueden participar en la gran lucha por comprender y dirigir sabiamente las leyes de la naturaleza y así lograr más de lo que es capaz la humanidad por sí sola, pero siempre de tal modo que los robots sirvan a los seres humanos.

Estas leyes de la robótica representan el primer intento declarado, sistemático y con un cierto impacto de crear una «ética de los robots», y fundan la primera aproximación que puede hacerse de este concepto, los sistemas éticos incorporados en los robots (10). Pero son también el punto de partida para otras interpretaciones y revisiones del concepto, como la ética de sus creadores, la ética de quienes usan los robots, y la ética acerca de cómo deberíamos tratarlos.

Ante todo, las Tres Leyes son un llamamiento a asumir las responsabilidades de nuestras creaciones. Una llamada de atención que sigue la tradición literaria instaurada en la modernidad por Mary Shelley en Frankenstein o El moderno Prometeo. Si el monstruo escapa y actúa mal, es solo porque su creador lo rechaza y se desentiende de sus acciones. En primera instancia, el doctor Victor es el responsable de los crímenes perpetrados por su criatura, pues actúa bajo la responsabilidad –o por la irresponsabilidad– de aquel.

De igual forma, los robots y las herramientas de inteligencia artificial no están desgajadas de la actividad de sus creadores, no son entes plenamente independientes ni autónomos. Obedecen a un programa impuesto y sus acciones son consecuencia del mismo. Como mucho, se podría establecer entre la actividad de la máquina y sus consecuencias, siguiendo la terminología de la atribución de la conducta de Fritz Heider, una relación de causalidad simple o justificabilidad (11): tiene capacidad de hacer lo que hace, pero lo que hace no está motivado ni es intencionado. Además, la causa de obrar como lo hace es situacional o externa; la máquina es obligada en todo momento por su propia condición programática –las máquinas no tienen nada parecido a una motivación, y las inclinaciones hacia una u otra opción conductual vienen derivadas de la intencionalidad de su creador en el momento de diseñarlas y dotarlas de una pauta de comportamiento–.

Que la ética de las Tres Leyes se despliegue en la ficción y no en un tratado o ensayo filosófico presenta la mayor de las ventajas: se inserta en la vida y obra de unos personajes que, salvando las distancias de la invención de sus mundos, la fantasía de los avances técnicos o el tiempo transcurrido desde que Asimov planteó su narrativa, recuerdan en todo a nuestra acción cotidiana y permiten desplegar una reflexión comparada.

Somos humanos, al fin y al cabo, nosotros y los que habitan en las páginas de estos libros, y aunque algunos autores han visto en la formulación de las leyes de la robótica un mero recurso literario con el que crear conflictos narrativos e ingeniosos entuertos que Asimov habría empleado para alimentar sus tramas criminales y legales, lo cierto es que el fundamento de este sistema ético guarda una enorme relación con la historia de la moral occidental: las Tres Leyes pueden resumirse en la máxima con la que Jesús condensa a su vez el espíritu de los diez mandamientos: «Amarás a tu prójimo como a ti mismo» (Mt 19, 16-19) (12). Como se verá más adelante, el propio Asimov pone estas mismas palabras en boca de uno de sus protagonistas más importantes, la robopsicóloga Susan Calvin, con objeto de explicar que el comportamiento de los robots está vinculado y se debe a un sistema ético común para toda la humanidad.

Al entroncar con nuestra tradición ética, la formulación de las Tres Leyes nos resulta natural, y su aplicación –la de algún tipo de restricción o reglamento ético sobre el uso de los robots y la inteligencia artificial– parece, a priori, una buena protección universal contra las amenazas, reales o imaginarias, que puedan suponer estas máquinas. Entre las más inminentes, y puede que la única efectiva en nuestros días procedente de la robótica –al margen de situaciones particulares generadas por un uso específico de herramientas concretas, como podría ser el caso los robots utilizados en medicina o actividades asistenciales–, su uso militar. Aquí contamos con numerosas y peligrosas aplicaciones, desde vehículos aéreos de combate no tripulados y la artillería pesada robotizada hasta los robots marinos y submarinos armados, pasando por los prototipos en un estado todavía embrionario de robots con orientación de diseño humanoide o animal con posibilidad de portar y manejar armas.

Para todo este armamento robotizado, y sobre todo para aquellos casos en los que el componente humano permanece «fuera de la cadena» de acontecimientos que llevan al uso final de estos recursos destructivos, la aplicación de unas leyes de la robótica y de la inteligencia artificial –sobre todo de un primer artículo como el formulado por Asimov–, supondría una contención alineada con el derecho internacional general ya existente: que ninguna vida humana esté «en manos» de un robot.

El militar no es el único debate al que se presta la aplicación de las Tres Leyes y la actuación de la robótica en el marco de los temas de fuerte actualidad de la discusión pública contemporánea. El caso más extremo y sobrecogedor, pero no por ello extraño a las inclinaciones sociales ya activadas en nuestros días, con el que Asimov nos muestra la injerencia de la Primera Ley en la comprensión de la vida humana en su conjunto tiene lugar en la novela El sol desnudo.

Klorissa Cantoro, la asistente genetista ayudante del oficial Rikaine Delmarre, encargada de supervisar a los embriones incubados artificialmente en el mundo espacial y robotizado de Solaria, responde con estas espeluznantes declaraciones a los interrogantes del detective Elijah Baley, terrícola en un mundo tecnológicamente

avanzado que no comprende:

—He observado que quienes cuidan a los embriones son robots. ¿No intervenían personalmente usted y el doctor Delmarre?

—A veces no teníamos más remedio que hacerlo, cuando las cosas se complicaban, por ejemplo, o cuando el feto presentaba defectos en su desarrollo. No se puede confiar en robots para tomar decisiones que pongan en peligro una vida humana.

Baley asintió:

—Se correría el riesgo de llegar a una conclusión equivocada y, como consecuencia, se podría perder una vida.

—Al contrario. Se correría el riesgo de sobrevalorar una vida y salvarla sin que lo mereciese. (SD, 169)

La Primera Ley obliga al robot a proteger y preservar la vida humana, por lo que no puede matar a un embrión. Tampoco el suicidio o la eutanasia serían posibles con un robot de por medio. Tal y como explica en Robots e Imperio otra de las grandes heroínas de Asimov, Gladia Solaria, cuando se refiere a la causa del escaso número de suicidios entre los espaciales: «Estamos rodeados de robots dedicados a mantenernos vivos. No hay modo de matarnos cuando nuestros activos y vigilantes robots están siempre a nuestro lado. Dudo de que alguno de nosotros lo intentara siquiera» (RI, 204).

De las conversaciones anteriores se extrae una parte que ya conocíamos: la protección con que la Primera Ley blinda la vida humana frente a la acción del robot. Al mismo tiempo, nos abren a

una nueva perspectiva, la de una sociedad en que los humanos tienen derecho a pronunciarse sobre el valor de la vida –la propia y la ajena–, juzgar cada una de sus manifestaciones sobre la base de una utilidad social para dictaminar si se merece o no ostentarla, y hacer con aquellas vidas de menor valor lo que se desee, incluso terminar con ellas. En estos pasajes, con independencia de la cosmovisión de cada cual acerca de la naturaleza y el propósito de la vida humana, es difícil no sentirse incómodo y tener la sensación de que los robots son poseedores de una humanidad superior a la propia exhibida por los humanos.

Para un robot, en situación ajena a un dilema fuerte en el que tuviera que decidir entre dos o más vidas, no hay distinción alguna entre personas y todas las vidas valen igual; no hay vidas «sobrevaloradas», ni vidas que no merezcan la pena ser vividas o salvadas.

Bien es cierto que el robot no actúa guiado por una decisión propia, sino obligado por norma, lo cual tiene poco o nada de meritorio en sí mismo, pero si comparamos su defensa de la vida, aun cuando esta es forzada, con la crueldad de acabar con una vida inocente e indefensa, entonces su acción gana muchos puntos, y nos pone ante la evidencia de la crueldad humana y la superioridad moral de los robots o, visto de otra manera más inquietante, de cómo la conciencia ética del ser humano podría caer hasta niveles inferiores a los de una máquina, aunque no por culpa de las máquinas.

De otra parte, la cuestión que veníamos tratando nos lleva a introducir la posibilidad del dilema, de esos momentos difíciles en los que los robots deben escoger de quién obedecer órdenes entre dos o más humanos que las emiten, o a quién proteger frente a dos o más humanos cuya integridad física está en peligro –un tipo de dilema del tranvía–. Lo cierto es que en este asunto nada de lo que nos cuenta Asimov acaba por ser definitivo.

Sí llamó la atención sobre su posibilidad y la dificultad de resolverlo –en el relato ¿Qué es el hombre?–, y señaló que, en caso de presentarse un dilema de esta índole, los robots primarían a los humanos más aptos según un baremo de «mentalidad, carácter y conocimiento»:

Cuando la Segunda Ley me exija obedecer a un ser humano, debo interpretar que he de obedecer a un ser humano que, por mentalidad, carácter y conocimiento, es apto para impartir esa orden. Y cuando hay más de un ser humano involucrado, al que, por mentalidad, carácter y conocimiento, sea más apto para impartir esa orden.

[...] si cada uno de los actos posibles supone daño para algunos seres humanos debo actuar de tal modo que garantice que el ser humano más apto, por mentalidad, carácter y conocimiento, sufra menor daño. (CC2, 642)

Sin embargo, nuestro autor nunca entró en detalles acerca de cómo se lleva a cabo esa evaluación de mentalidad, carácter y conocimiento y qué cuenta como tal. De hecho, en las ocasiones en que sus tramas propician la llegada de un dilema de esta índole, en las que la acción del robot en uno u otro sentido se sabe decisiva, ni siquiera la propia máquina es capaz de determinar a ciencia cierta y con toda seguridad cuál será la mejor de las posibilidades.

Esta indefinición podría ser simbólica, y hasta una declaración de intenciones, en cuanto a que con ella Asimov evita convertir la justicia y los derechos en una operación matemática, en un problema al que podamos encontrarle una solución exacta atribuyendo a cada bien humano una medida simple a integrar sin ningún tipo de consideración cualitativa en un mero cálculo –la existencia de unas leyes de la robótica reforzaría esta importancia de los principios sobre los cálculos–.

Con este planteamiento, no solo la respuesta utilitarista quedaría descartada (13), sino que se nos invitaría a dar a cada dilema una solución que tenga en cuenta las cualidades humanas involucradas en cada caso, lo cual incluiría, a su vez, la necesidad de discernir a nivel personal y propiciar el diálogo y la discusión pública en comunidad, al modo en que las sociedades demuestran buscar la

vida buena y virtuosa que les da a ellas mismas la cualidad de humanas.

La Ley Cero

La fe de Asimov en las Tres Leyes de la robótica es casi absoluta. Tanto que, para nuestro escritor, podemos fiarnos mucho más de un robot que de un ser humano. En el escenario de estas leyes, la desconfianza irracional vertida sobre los robots no sería más que «una enfermedad de la Humanidad, de la cual todavía no se ha curado» (VR, 303).

Decimos que es una fe «casi» absoluta, pues encontramos en su narrativa la excepción hecha por la denominada «Ley Cero» de la robótica, presentada en Robots e Imperio, el último volumen del ciclo de los robots, por el robot Daneel Olivaw:

Hay una ley que es superior a la Primera Ley. «Un robot no puede lastimar a la humanidad o, por falta de acción, permitir que la humanidad sufra daños.» La considero ahora la ley Zeroth [Ley Cero] de la Robótica. La Primera Ley debería decir: «Un robot no debe dañar a un ser humano, o permitir, por inacción, que el ser humano sufra algún daño, a menos que tal acción viole la ley Zeroth de la Robótica». (RI, 380)

Este cambio de perspectiva de una protección de la persona individual y concreta –que Asimov apuntaló en la práctica totalidad de su narrativa sobre robots– a una defensa de la humanidad por encima del individuo –que Asimov solo desarrolla por entero hacia el final de su carrera literaria– puede estar motivado por una mudanza en su manera de pensar al respecto, o bien responde a un recurso narrativo para terminar de unir la saga de los robots con la saga de Fundación, pues la Ley Cero guarda una relación innegable con la naturaleza y principios de la psicohistoria (14) y, al fin y al cabo, Robots e Imperio fue un eslabón clave, junto con Fundación y

Tierra para conseguir que ambas series literarias quedaran conectadas. De hecho, el final de esta última novela apunta de nuevo a las modificaciones realizadas en las Tres Leyes de la robótica para redondear el sentido final de la trama.

En cualquier caso, tras las palabras de R. Daneel Olivaw (15), Asimov recurre a otro robot, Giskard Reventlov, para matizar su punto de vista y llamar la atención sobre una cuestión ética importante que enlaza de manera directa con buena parte de la historia del siglo XX:

No puedo aceptar la ley Zeroth, amigo Daneel —respondió Giskard lentamente—. Sabes que he leído mucho sobre la historia humana. En ella he descubierto grandes crímenes cometidos por seres humanos y la excusa era siempre que estaban justificados por las necesidades de la tribu, del estado o, incluso, de la humanidad. Es precisamente por ser la humanidad una abstracción por lo que se la utiliza libremente para justificar cualquier cosa y, por tanto, tu ley Zeroth es inconveniente. (RI, 381)

Milenios más tarde, en Fundación y Tierra, un superviviente robot Olivaw, interrogado por Golan Trevize acerca de la difícil aplicación de la Ley Cero, volverá a expresar el mismo temor, aunque en forma de evidencia:

—¿Cómo deciden lo que es o no es perjudicial para la Humanidad en su conjunto?

—Ahí estriba el problema, señor —dijo Daneel—. En teoría, la Ley Cero era la solución de nuestras dudas. En la práctica, nunca podíamos decidir. El ser humano es un objeto concreto. Los daños a una persona pueden ser calculados y juzgados. Pero la Humanidad es una abstracción. ¿Cómo resolver esta dificultad? (FT, 510)

La excepción de la Ley Cero, aun cuando raya en el límite del fenómeno fandom y sus debates acerca de si resulta un buen recurso o no para unir dos sagas en principio separadas, tiene su interés por cuanto que nos alerta acerca del peligro de anteponer el cumplimiento de un programa ideológico al bienestar real de las personas, sometiendo las cualidades de la persona a la uniformidad de solo una idea.

Un defecto al que, decía el historiador Paul Johnson, suelen acostumbrarse los intelectuales y quienes les prestan oídos –pero no solo los intelectuales, examínese cada cual– cuando olvidan «que las personas importan más que los conceptos y deben ser colocadas en primer lugar. El peor de todos los despotismos es la tiranía desalmada de las ideas» (16).

Con la tentativa de esta ley, Asimov vuelve a recordarnos que la humanidad no puede reducirse a un código, y que los problemas acerca de lo que resulta más justo no son una operación matemática que pueda resolverse mediante un sistema algorítmico cerrado. La historia –toda la historia, y en particular la más reciente– nos ha dado muestras de los desastres y terrores derivados de querer someternos a una serie de variables que deben componer una fórmula con un único resultado.

La riqueza del espíritu humano desborda todo intento de cálculo. No importa lo exhaustivo o perfecto que este cálculo aspire a ser en su representación de la realidad. Esta es, en primer o en último término inabarcable e inasible, y tampoco es susceptible de ser toda ella matematizada. La humanidad presenta a cada paso que da, en lo grande y en lo pequeño, imponderables que, en el fondo, en toda su extensión resultan inmanejables por el procedimiento de inferencia inductiva de una máquina, por mucha fuerza de computación con que la dotemos y por mucho tiempo de aprendizaje profundo –deep learning– que incorpore.

Lidiar con distintas sensibilidades, dialogar y mediar de manera constante, encontrarse y entenderse con los otros distintos, en definitiva, convivir con todas nuestras particularidades y diferencias, por muy difícil que sea la tarea encomendada, debería

seguir siendo la forma de organizarnos en sociedad. Es lo que nos hace humanos, ese es el estándar de mejora continua que necesitamos, y no el de una máquina uniformadora.

Reaccionar ante el futuro que viene seguirá siendo tarea del ser humano si no queremos repetir los errores del pasado sometiéndolo o apostándolo todo, una vez más, a una sola idea con apariencia de perfección maquinal, pero en verdad incompleta e insuficiente, pues no contempla ni ha lugar a las imperfecciones que nos constituyen.

El complejo de Frankenstein

Más allá de la reflexión y relación de las leyes de la robótica con la ética –la humana, pero también la ética de los robots– que ya hemos abordado antes, resta por explorar una aplicación práctica de estas normas que también resulta interesante. Ya señalamos la correspondencia de estos principios con la tradición moral de Occidente, por lo que no debería extrañarnos la escasa novedad que, en realidad, y exentas de su aura ficcional, presentan estas leyes con respecto a otras normas que la humanidad ha utilizado para otros artefactos tecnológicos a lo largo y ancho de su historia.

Para llamar la atención sobre este punto, Asimov despliega en paralelo a las Tres Leyes de la robótica las denominadas Tres Leyes de las herramientas:

1. Una herramienta debe ser segura para poder

emplearla.

2. Una herramienta debe llevar a cabo su función, siempre y cuando la realice con seguridad.

3. Una herramienta debe permanecer intacta durante su uso, a menos que se requiera su destrucción por motivos de seguridad, o que su destrucción constituya una parte de sus funciones.

Para Asimov, por tanto, el robot no es más que otra herramienta,

resultado de los avances en el conocimiento y en la técnica de la humanidad. Resultado y avance por el cual debemos sentirnos orgullosos, porque es fruto de grandes esfuerzos humanos, presupuestarios y técnicos. Este es el espíritu de admiración y grandeza que se deriva del parecer de Lucky Starr, héroe protagonista de la Federación Terrestre en la novela Los anillos de Saturno (17) incluso cuando lo que valora son los robots de los enemigos de la Confederación Siriana:

Esos robots son una conquista del hombre. Los seres que la llevaron a cabo eran sirianos, en efecto, pero son también seres humanos, y todos los demás miembros de la especie pueden sentirse orgullosos de esta victoria. Si nosotros tememos los frutos de dicha victoria, realicémosla también, y superémosla incluso. Pero de nada sirve negarles el mérito de lo que han conseguido. (AS, 90)

Como toda herramienta, el robot puede ser utilizado para bien o para mal por los seres humanos. En último término, como todo mal derivado del uso de cualquier herramienta, la causa o culpa del uso que se hace de una herramienta reside en quienes la utilizan.

Las personas disfrutan de libre albedrío, condición de la que el robot carece –de nuevo, como cualquier otra herramienta–, sometido por defecto de fábrica a una serie de condiciones de programación –en el caso de la literatura de Asimov, a las Tres Leyes de la robótica–. La persona puede caer en falta –ceder ante las tentaciones y su propia debilidad es, al menos, una opción para el ser humano–, mientras que el robot está obligado a cumplir con la programación dada, nunca traicionará las instrucciones dadas.

El robot no puede, sin más, por ciencia infusa, mediante el ejercicio de un libre albedrío que no posee o la activación de unas instrucciones que no están en su programación escapar de esa red de normas preestablecidas, como muchos singularistas, futurólogos o expertos en computación entregados a la elucubración pretenden cuando hablan del salto que la inteligencia artificial estrecha dará a una inteligencia artificial general, y de ahí a una superinteligencia.

¿Cómo? Las explicaciones que siguen a esta advertencia son siempre ciencia ficción, más propias de la novela que del ensayo.

En todo caso, si se produjera una contradicción entre dos órdenes de similar prioridad o una orden incomprensible o contradictoria con las instrucciones de la máquina se produciría lo que Asimov dio en llamar para sus cerebros positrónicos un bloqueo robótico o «robloqueo», como se apunta con creatividad etimológica en Robots del amanecer. En el peor de los casos, un fallo sistémico del robot ocasionado por desperfectos en su maquinaria debido a causas ambientales podría dar lugar a un accidente –como el robot de la Confederación Siriana de El gran sol de Mercurio, que casi acaba con la vida de Lucky Starr, pero por accidente–. Como toda herramienta, los robots también pueden fallar y ser motivo de desgracia, aunque no intencionada. Si no hay libertad, cómo podría haber intención.

La persona puede caer en falta debido a su libre albedrío, pero por ese mismo motivo, el ser humano es también, en el ejercicio de su libertad, capaz de grandes bienes. Por el contrario, el robot obra bien de acuerdo con una norma preestablecida. No puede obrar de otra manera, está obligado a hacer el bien –lo cual, por otra parte, le resta mérito–.

De ahí la famosa aseveración del personaje y padre de familia George Weston en el relato Robbie –célebre por cuantas veces ha sido recuperada y citada por el propio Asimov para subrayar esta idea, y por ser parte de su primera incursión en la narrativa sobre robots– para tranquilizar a su mujer con respecto al comportamiento del robot Robbie, que trabaja para la familia como niñero (18): «Toda su “mentalidad” ha sido creada para este propósito. Sencillamente no puede evitar ser leal, encantador y amable. Es una máquina… hecha así» (VR, 71).

De ahí también la incomprensión del robot Daneel Olivaw cuando el detective Elijah Baley le describe en Bóvedas de acero la influencia positiva que tiene la religión en el comportamiento de las personas. El robot no entiende cómo una ética expresada en un libro –la Biblia–, pero no programada, no codificada, que no ha sido grabada en el cerebro de las personas puede ser una guía hacia el bien, dado que él, como robot, solo conoce la privación de libertad

resultado de que sus acciones estén todas ellas sujetas sin excepción a las Tres Leyes de la robótica.

Sin embargo, a pesar de la evidencia racional de los límites que impone la programación –sean las leyes de la robótica en la ficción o los sistemas de computación utilizados en nuestras máquinas–, a pesar de sus múltiples garantías de cumplimiento, el ser humano se ve invadido por emociones que le llevan a temer al robot, en lo que Asimov denomina en sus relatos el «complejo de Frankenstein»: «La Humanidad puede conocer la existencia de las Tres Leyes a un nivel intelectual y, sin embargo, tener un miedo por completo desarraigable y desconfiar de los robots a un nivel emocional. Si desean inventar un término para esto, lo podrían llamar "Complejo de Frankenstein"» (VR, 470).

En este sentido, cualquier problema que pueda derivarse de la acción de un robot tiene en su origen las malas intenciones de la persona o el colectivo de personas que miente y confunde al robot para favorecer sus intereses.

En todos los casos, los robopsicólogos que protagonizan los relatos de Asimov son capaces de deshacer el entuerto y poner en evidencia la falta cometida por la persona, que no por el robot, sirviéndose de la predisposición paradójica del robot hacia el bien: los robots, al contrario que los humanos, nunca se defienden a sí mismos y a costa de otros; el ser humano sí, y en su egoísmo espera que el robot haga lo mismo. Nunca ocurre así, y es «fácil» descubrir en falta a la persona por esa inclinación a esperar de otros –los robots– su misma actitud mezquina.

Ocurre de este modo en el relato Ha desaparecido un robot (VR). A unos pocos de los modelos Nestors (NS-2) les es sustraído en secreto un fragmento de la Primera Ley. Se conserva la parte inicial –«ningún robot puede hacer daño a un ser humano»–, pero se elimina de su sistema la segunda –«o, por medio de la inacción, [ningún robot puede] permitir que un ser humano sea lesionado»–.

El objetivo: que ningún robot apartara a los trabajadores de la Base Hiper, dedicados a las tareas de desarrollo del Viaje Hiperatómico, de su puesto de trabajo, a causa del posible daño de los rayos gamma a los que en ocasiones podían estar expuestos sin, en teoría,

riesgo real para su salud. El robot acompañante, fruto de su lectura de peligro por la exposición a los rayos, interpretaba la situación como un riesgo potencial, sacaba a rastras a su protegido y le ponía a resguardo, interrumpiendo el trabajo y ralentizando los avances.

En la narración, el desequilibrio generado en el cerebro positrónico por la modificación de la Primera Ley dará lugar al intento de fuga de uno de los robots afectados, poniendo en jaque el secreto empresarial y de Estado de la alteración y llevará a cabo una investigación capitaneada por Susan Calvin.

La reflexión aquí no es tanto acerca del peligro de dicho desequilibrio para la integridad física de una persona, aunque también. Por ejemplo, un robot afectado por este cambio podría lanzar un objeto pesado sobre un humano sabiendo que puede coger el objeto al vuelo antes de que impacte en la persona –no se incumpliría la parte todavía activa de la Primera Ley–; sin embargo, una vez el objeto dejara de tener contacto con el robot, este podría permitir que el objeto impactase con el objetivo sin producirse ningún incumplimiento, dado que la segunda parte, que protegía por inacción, ya no consta como obligación.

Este es un peligro a tener en cuenta, y pone de manifiesto el siempre temido atrevimiento humano, y la culpa de los ingenieros –que no del robot, que solo cumple con lo que se ha grabado en su sistema– al efectuar un cambio de consecuencias fatídicas para las personas.

Sin embargo, la verdadera reflexión de fondo viene de tomar conciencia de la intención original al obrar el cambio en la Primera Ley. Los responsables del proyecto del Viaje Hiperatómico pusieron el cumplimiento de la misión, y la maximización del rendimiento y del beneficio empresarial por encima de la seguridad de sus empleados –de los que se hace saber al comienzo del relato que más de uno cayó herido por los rayos gamma a pesar de que «eran seguros»–. Por lo tanto, es la lógica retorcida de una economía que sitúa las ganancias en el centro del modelo, y no la seguridad, la integridad física y el bienestar de las personas, lo que propicia la entrada de una orden malsana en el sistema del robot, y hace de este un ente peligroso para los tripulantes de la Base Hiper y, por ende, para la humanidad.

Con todo, no son los robots la gran amenaza para el mundo. Somos nosotros, responsables del funcionamiento de esos robots a través de las reglas que insertamos en ellos, de las alteraciones de esas reglas con un fin dañino para uno mismo y para los demás, o de las extrapolaciones en extremo negativas e imaginarias con que dotamos a estas máquinas y que tienen como único resultado la parálisis y el miedo. Si acaso somos nosotros, con nuestras ínfulas de infalibles creadores, como en todo tiempo y lugar desde que el hombre es hombre, sea cual sea, entonces y ahora la herramienta fetiche para la expresión de nuestro dominio técnico sobre la realidad y la manifestación de nuestra naturaleza imperfecta, somos nosotros la gran amenaza para el mundo.

▪

7 En la narrativa de Asimov es habitual que los números de serie de los robots y sus nombres sigan el mismo patrón fonético. Así, las siglas «RB» de RB-34 se pronuncian en inglés «ar-bi», muy similar a Herbie. Otras veces, el nombre del robot busca reproducir o asemejarse a los fonemas del número de serie: el robot JG-10 es George Diez.

8 Hermanos Čapek, R. U. R. y El juego de los insectos.

9 Martínez, La ciencia ficción de Isaac Asimov, p. 119.

10 Asaro, «What Should We Want From a Robot Ethic?», International Review of Information Ethics, 6.

11 Heider, The Psychology of Interpersonal Relations.

12 No hubo en Asimov un impacto reseñable del judaísmo que, por sus orígenes, podría suponérsele como un componente esencial de su cosmovisión del mundo. La dimensión religiosa no se manifestó en su vida. Fue siempre ateo y racionalista, y en lo personal nunca se ligó a ninguna corriente de pensamiento que diera importancia a realidad sobrenatural alguna. La explicación material de la existencia fue siempre su guía para entenderlo todo.

Para describir su actitud ante la realidad, cabría hacer suya la cosmovisión de uno de los personajes secundarios de Fundación y Tierra, el historiador Vasil Deniador del planeta Comporellon, perteneciente a un grupo de intelectuales conocido como los Escépticos: «Solo acepto aquello que las pruebas lógicas me obligan a aceptar y aún mantengo en suspenso dicha aceptación hasta que otras pruebas me lo confirmen» (FT, 151).

Sin embargo, su ateísmo no fue óbice para que mostrara un genuino interés intelectual por la religión y el máximo respeto hacia una fe no compartida.

13 En el modo en que lo exponemos con anterioridad y recoge con mayor detalle y sistematicidad Sandel, Justicia. ¿Hacemos lo que debemos?

14 En el universo de ficción de Asimov, una ciencia que combina la psicología, la historia y las matemáticas para predecir el futuro de la civilización humana.

15 La «R.» se utiliza en las sociedades robóticas como abreviatura de «robot», al modo en que para nosotros «Sr.» o «Sra.» son abreviaturas de «señor» y «señora».

16 Johnson, Intelectuales, p. 351.

17 En realidad, es el gran héroe de toda una saga de seis novelas, la incursión más conocida de Asimov en la literatura juvenil (la escribió bajo el pseudónimo de Paul French).

18 Asimov vaticina en tiempos tan tempranos como 1940, año de publicación de este relato, una de las tres funciones básicas que tendrán en nuestros días los robots y los sistemas de inteligencia artificial en conexión con la dimensión afectiva y emocional: robots niñeras, robots cuidadores de ancianos y discapacitados, y robots compañeros sexuales –este último rol, que más adelante será objeto de comentario junto con el resto, también fue tratado con antelación en Los robots del amanecer, en la relación afectivo-sexual de Gladia Delmarre con el robot humaniforme Jander Panell.

¿Qué es lo humano?

En este libro nos desmarcamos del discurso y posicionamiento más habitual sobre la amenaza que representan los robots y el riesgo de rebelión de las máquinas que las llevarían a intentar desbancar, sustituir o incluso destruir a la humanidad, en un eterno recuerdo a la acostumbrada epifanía de la rebelión del hombre contra el Creador.

Al respecto, dice Asimov en el ensayo breve Mis robots, un texto que trata acerca de sus creaciones robóticas en la ficción, algo que podríamos tener en cuenta como factible en la construcción y programación de estas máquinas en el panorama de desarrollo tecnológico actual: «Los construí de manera tal que no pudieran atentar contra su creador, y habiéndolos despojado desde su inicio de toda posibilidad de causar daño, me encontré libre de otorgarles aptitudes más racionales» (VR, 521-522).

Desmarcarse de esta línea narrativa tan concurrida por otros autores permite poner en su sitio al hombre y a la máquina, atribuir a cada uno sus características más definitorias y prometedoras, e iniciar con el lector una reflexión sobre la distinta naturaleza de ambos entes.

A través del comentario de fondo de las potencialidades y limitaciones de los robots para dar respuesta a las demandas de la vida, Asimov perfila a su vez las potencialidades y limitaciones de las personas. La persona explica al robot, y el robot explica a la persona. Estas potencialidades y limitaciones se refieren a todas las dimensiones humanas –conducta, afectividad y cognición–, incluyendo, por supuesto, la reflexión sobre la gran pregunta de actualidad acerca de las posibles implicaciones futuras de una inteligencia artificial cada vez más avanzada.

Si la inteligencia artificial, en palabras de la especialista Margaret

A. Boden, «tiene por objeto que los ordenadores hagan la misma clase de cosas que puede hacer la mente» (19), cabe preguntarse si su objeto es posible y realizable –no solo desde el punto de vista tecnológico–, es decir, si la mente alberga o no un reducto de propiedades y capacidades singulares reducibles a la categoría de «“máquina virtual” insertada en el cerebro»; si la reproducción de dichas propiedades y capacidades resulta alcanzable o no más allá de una pobre imitación o, en el mejor de los casos, una imitación imperfecta; o si la mente «contiene» o «trabaja» o no con realidades imposibles de plasmar en cualquiera de los sistemas de representación del conocimiento disponibles para configurar la inteligencia artificial.

Todas estas y otras muchas inquietudes con una orientación muy similar son habituales en el debate permanentemente abierto en nuestros días sobre la inteligencia artificial, la cibernética y la robótica. En las ficciones de Asimov, si no una respuesta técnica y especializada, sí encontramos al menos apuntado, entre sus diálogos o como parte de los mismos, un punto de vista ponderado y accesible a todos los lectores desde el cual iniciar y orientar la reflexión personal para desplegar una respuesta.

Al respecto encontramos en su obra, repetida de muy distintas maneras, más o menos directas o explícitas, más o menos literarias o realistas, una idea fundamental a partir de la cual articulamos la reflexión aquí: que humano y robot son entes distintos, que no son lo mismo y que no existe método o técnica capaz de que ambos lleguen a ser una misma cosa.

No importan la cantidad y cualidad de los avances tecnológicos que puedan acontecer en el robot en cuanto a su capacidad de procesamiento, sus funcionalidades y su apariencia humanoide; tampoco las «mejoras» que puedan introducir los humanos en sus cuerpos para disfrutar de una potencia física superior, de una resistencia corporal sin parangón o de una salud reforzada. Humano y robot son cosas distintas. Las personas presentan elementos diferenciales de grado que las máquinas no pueden siquiera aspirar a imitar. Al fin y al cabo, el humano vive, la máquina no, la máquina funciona: «—Tú estás vivo, Daneel, ¿verdad? [pregunta Elijah Baley al robot Daneel Olivaw.] Daneel contestó, lentamente y

con énfasis: —¡Yo funciono!» (RA, 42).

Pero más allá de esta máxima generalizadora, quizá sea el relato El hombre bicentenario el más importante y célebre, acerca del límite infranqueable entre ambos entes, expresado en este caso desde la perspectiva del robot.

En el transcurso de la trama y de manera progresiva, el protagonista, el robot Andrew va avanzando en derechos con anterioridad solo reservados a las personas: consigue la libertad, por lo que deja de estar sometido al control de sus dueños y empieza a trabajar para sí mismo; tiene derecho a la propiedad privada, del que obtiene su propia casa y cuenta bancaria, con ahorros que le permiten disfrutar de la acumulación de bienes; comienza a vestir como quiere y se somete a diversas cirugías, cada una de ellas más arriesgada que la anterior, para conseguir un organismo lo más humano posible. La última de dichas operaciones le llevará a reducir su esperanza de vida para ganar la condición de «mortal».

Con todo, a pesar de todos sus esfuerzos, de todos sus «sufrimientos», del drama que supone su final para las personas que le tienen cariño y se solidarizan con su causa, Andrew nunca llegará a ser aceptado como hombre. Porque no lo es. Andrew es un ente sometido a los principios de la robótica y de un sistema programado, limitado a capacidades que emulan lo humano. La emulación se ha logrado con gran perfección y detalle, sin duda. Pero, en definitiva, no trasciende a su condición de mera imitación a la que está condenado por la naturaleza de su propia configuración. Andrew es creado, pero no nacido. Andrew es un robot.

Quizá podríamos terminar de observar su caso a ojos de Golan Trevize –el héroe de Los límites de la Fundación y Fundación y Tierra–, cuando, juzgando un caso similar, el del robot humanoide Daneel Olivaw, dice: «No podía ser percibido como una simple mente robótica. Pero tampoco podía ser percibido como humano» (FT, 520).

La apreciación de Trevize sigue unos derroteros muy similares a las impresiones de quienes asistimos a las sorprendentes competencias y resultados de las herramientas de inteligencia artificial de uso más

extendido, incluso cuando el objeto de valoración no es ni mucho menos un robot humanoide con cerebro positrónico, sino algo mucho menos potente y ambicioso como una de las inteligencias artificiales generativas cuyo uso cunde hoy entre millones de usuarios, con funciones en realidad acotadas a un área de desarrollo muy concreta y no poco limitadas en cuanto a su alcance real.

Cuando vemos las espectaculares imágenes e ilustraciones que pueden generar aplicaciones como Adobe Firefly o Leonardo AI con tan solo recibir una breve descripción escrita de lo que se quiere obtener, o el detallado realismo de los vídeos de Sora, con planos y movimientos de cámara de gran profesionalidad y una habilidad inédita de emular distintas ópticas y ofrecer recreaciones históricas, nuestra primera reacción es pensar, en la línea del personaje de Asimov, que estamos ante algo más que una máquina, pues las imágenes, vídeos o textos que ofrecen estas herramientas están a un nivel de ejecución profesional, es decir, a un nivel muy por encima de las habilidades y del material de creación del que disponen la gran mayoría de los usuarios que las utilizan. En refuerzo de esta idea y de la sorpresa que conllevan las respuestas de estas máquinas, no hay que perder de vista que obtenemos estos materiales audiovisuales de gran calidad en muchísimo menos tiempo –una media de 30 segundos– del que habría necesitado un profesional para llevarlos a cabo.

En realidad, todas estas herramientas tienen serias limitaciones de funcionamiento y presentan múltiples y evidentes errores en sus resultados, como cualquiera que las utilice con una cierta regularidad –y no se ciña al vídeo de muestra de la corporación de turno para emitir su juicio– habrá podido comprobar –y sufrir–. Fallos de bulto que un profesional nunca cometería y un principiante nunca los presentaría como buenos porque son apreciables a simple vista –por el ojo humano, al menos–. Pero incluso cuando su funcionamiento fuera irreprochable y sus resultados sin falla sumamente impresionantes, estos seguirían sin ser humanos.

Más allá de un uso instrumental y superficial de estos materiales audiovisuales o artísticos, esta condición no es ni mucho menos accesoria. Al margen de la curiosidad pasajera de ver una película

realizada por máquinas, un concierto interpretado por robots, o un cuadro pintado por sistemas de inteligencia artificial, ¿qué interés real pueden tener esas obras de arte? Detrás de una película así no cabría especular acerca del peso que han tenido las experiencias vitales del director en el metraje, ni sobrecogerse por la pasión espontánea e irrepetible que los músicos han puesto en su interpretación, ni deslumbrarse con la genialidad del pintor que lleva a cabo la proeza de plasmar la condición humana en un lienzo. La máquina carece de todas las cualidades que hacen especiales a las personas, las mismas que hacen especiales sus obras.

Asimov se permitió bromear con esta idea en Navidades sin Rodney (VR). La mujer del protagonista de este relato quiere proporcionar unas pequeñas vacaciones a su robot doméstico Rodney, pues ha acabado por humanizar a la máquina después de tanto tiempo unidos en el día a día del hogar; ante la negativa de su marido, insiste: «Vamos, Howard —respondió—. Rodney no es un frigorífico ni un esterilizador. Es una persona». Howard responde tajante: «No es una persona. Es un robot. No desearía unas vacaciones» (VR, 453-454).

En broma, pero también en serio, Asimov proporciona con esta respuesta otra perspectiva interesante de las diferencias insalvables entre humanos y robots: la ausencia de deseos, ambiciones y objetivos propios en las máquinas, reducidas a un limitadísimo marco normativo fijado por terceros y compuesto de condiciones algorítmicas que están obligadas a cumplir. ¿Puede tener un robot un objetivo? Sí, el que el programador haya especificado como tal, marcando un cierto resultado con la etiqueta de «éxito» y dejando fuera el resto como «fracaso». La máquina, a partir de este criterio, mediada por las instrucciones iniciales, podrá lanzar inferencias para llegar a resultados similares. ¿La máquina «comprende» o «razona» en algún momento sobre la validez de estas funciones y su conveniencia? ¿Expresa sus preferencias por una u otra vía de actuación? Desde luego que no.

La incorporación de tales deseos, motivaciones o cualquier otro tipo de agencia, libertad o voluntad propia a una inteligencia artificial pasaría por computerizar –reducir a una operación matemática– valores y emociones humanas para las que ni siquiera la filosofía o

la psicología tienen todas las respuestas, y en cuya complejidad y matices los propios humanos nos vemos enzarzados, cuando no perdidos a diario. Trasladar hasta el más simple de estos deseos, emociones y valores al código de una función de utilidad es tarea irrealizable. Aun cuando dispusiéramos de toda la fuerza de computación que pudiéramos imaginar, seguimos sin poseer el conocimiento completo de estas manifestaciones, por lo que estaríamos programando algo en definitiva incompleto.

Cualquier otro método de réplica o generación de los mismos más allá de la programación estándar y el aprendizaje profundo –métodos proyectados desde el plano teórico como la selección evolutiva, el aprendizaje por refuerzo, la acumulación por asociación, el andamiaje emocional, la modulación de emulaciones, entre otros (20)– no son más que ciencia ficción: cualquiera con habilidades narrativas y unas condiciones aceptables de imaginación podría especular acerca de ellos e inventar los suyos propios.

«Un acto humano conocido como perdonar»

El deseo, así como otras características psíquicas y emocionales como la curiosidad y la intuición sirven a Asimov para marcar una primera diferencia cualitativa entre el hombre y la máquina. En ninguna de sus apuestas narrativas es más clara esta intención de ir dejando un rastro progresivo de diferencias entre ambos como en las novelas protagonizadas por la memorable pareja de investigadores Elijah Baley –humano terrícola– y R. Daneel Olivaw –el exclusivo robot humanoide fabricado en la sociedad espacial de Aurora–.

Al robot Daneel nunca le mueve el anhelo de iniciar un camino propio en pos del desarrollo personal. La curiosidad por poseer nuevos conocimientos con el objetivo de rellenar las lagunas del saber y disponer de nueva información desde la cual llegar más lejos en el proceso deductivo, o la necesidad imperiosa, urgente, y en muchos casos inútil de llegar hasta el final para terminar de conocerlo todo acerca de una cuestión que nos interpela, pero de la que nos falta saber una parte es una sensación que el robot no es capaz de entender, y todo lo más que puede hacer Elijah es intentar definirla, aunque no sea aprehensible o interiorizable por la máquina:

—¿No tienes curiosidad personal, Daneel? —dijo—. Te has llamado detective. ¿Sabes lo que eso requiere? ¿Entiendes que una investigación es más que una tarea? Es un desafío. Tu mente se enfrenta a la del delincuente. Es un choque de intelectos. ¿Puedes abandonar la batalla y admitir la derrota?

—Si continuarla no sirve a ningún fin valioso, desde luego.

—¿No sentirías ninguna pérdida? ¿Ninguna duda? ¿No habría siquiera una mota de insatisfacción? ¿De curiosidad frustrada?

Las esperanzas de Baley, que no eran grandes de entrada, se empequeñecieron según hablaba. La palabra «curiosidad», repetida dos veces, le recordó sus propios comentarios a Francis Clousarr cuatro horas antes. Había sabido muy bien las cualidades que distinguían al hombre de la máquina. La curiosidad tenía que ser una de ellas. Un gatito de seis semanas era curioso, pero ¿cómo podía ser curiosa una máquina, por muy humanoide que fuera?

R. Daneel se hizo eco de estos pensamientos al decir:

—¿Qué quieres decir con curiosidad?

Baley se preparó para hacer lo que pudiera.

—La curiosidad es el nombre que damos al deseo de aumentar el conocimiento que uno tiene.

—Ese deseo existe en mí, cuando el aumento del conocimiento es necesario para cumplir una determinada tarea.

—Sí —dijo Baley, sarcásticamente—, como cuando preguntas sobre las lentes de contacto de Bentley para saber más sobre las peculiares costumbres de la Tierra.

—Exactamente —dijo R. Daneel, sin dar signos de haber notado el sarcasmo—. El aumento del conocimiento sin propósito, sin embargo, que es a lo que creo que en realidad te refieres con la palabra curiosidad, es simplemente poco eficiente. He sido diseñado para evitar la ineficiencia. (BA, 244-245)

El robot ha sido creado bajo la premisa de la eficiencia y sigue la máxima del utilitarismo habitual en los sistemas de cálculo: descartará el conocimiento que no rinda un beneficio inmediato por supuestamente inútil. Lo cierto, sin embargo, es que en buena medida los casos a los que se enfrentan ambos detectives son solucionados, en última instancia, gracias al apoyo circunstancial de ese excedente de saber en el que se empeña el detective humano Elijah. Esa información o conocimiento que no parece ofrecer un rendimiento compensatorio inicial, pero sí acaba teniendo una

importancia notable al final, de forma que ese anhelo tan humano por conocer no sirve para nada, pero al mismo tiempo sirve para todo –en la provocativa línea de «la utilidad de lo inútil» de la que nos hablaba el profesor y escritor italiano Nuccio Ordine–.

De una forma parecida, parece darse una distinción entre máquina y ser humano con la oposición de lógica (cerebro positrónico) frente a razonamiento (mente humana), entendiendo esta última como una capacidad que va más allá de aquella al hacer uso de facultades complementarias que la enriquecen, como la reflexión, la imaginación o el presentimiento.

En la resolución de los casos policiales planteados en el ciclo de los robots, casos de asesinatos en los que participan la pareja Elijah y Daneel como investigadores, son muchas las ocasiones en que el discurso lógico maquinal del robot Daneel Olivaw, con su exhaustivo repaso a toda información disponible sobre el caso plantea posibles soluciones para el crimen, pero este proceso de conexión lógica de los diferentes eventos acaecidos no alcanza para dar con una solución definitiva. Lo que alcanza a solucionar el crimen es el razonamiento del detective Elijah, quien, movido por la intuición o por un momento eureka fruto de un elemento de apariencia casual, logra desencadenar asociaciones que conducen hasta la resolución del caso (21).

De esto último es incapaz el robot, no digamos ya de la explotación de las debilidades del interrogado, requiebros del lenguaje o engaños situacionales, habilidades todas ellas dependientes de sutilezas indetectables por la máquina, y argucias de las que ningún robot es capaz.

Primero, porque la mentira supone traicionar el bien absoluto e inquebrantable al que están sujetos sin excepción. Segundo, y más importante, porque la intuición no forma parte de un hilo de razonamientos conscientes, por lo que no se puede dividir en pasos lógicos y, por tanto, no es programable y no puede entrar a formar parte de la inteligencia de la máquina.

Atender a la posibilidad o imposibilidad de computerizar la intuición es sumamente relevante, pues se encuentra en la base de la discusión sobre la posibilidad o imposibilidad final de la

inteligencia artificial. Esta diatriba se manifestó en fecha tan temprana como 1938 en la tesis doctoral de Alan Turing, Systems of Logic Based on Ordinals.

En aquel momento, Turing reconoció que la intuición no era matematizable, dando razón a los teoremas de incompletitud del matemático Kurt Gödel. Sin embargo, a partir de los años cincuenta, en concreto a partir del célebre artículo Maquinaria computacional e inteligencia (1950), Turing trabajó en la idea de que las operaciones matemáticas complejas destinadas a desarrollar y potenciar el ingenio podían compensar esa falta de intuición, haciendo de esta una competencia en último término innecesaria, siempre y cuando se entendiera la inteligencia como la acabó entendiendo Turing y la entendieron a partir de entonces todos los entusiastas mesiánicos de la inteligencia artificial: algo que resuelve problemas.

No sabemos a ciencia cierta si Asimov dio tamaña importancia a la intuición en su narrativa, o eligió plantear su ciclo de los robots en el escenario de una trama policial –de nuevo la resolución de un problema, de un acertijo– a expensas de o en respuesta a la aportación de Turing y de su reducción del alcance y dimensiones de la inteligencia, pero acertó poniendo en primer plano de su literatura sobre robots esta capacidad humana, pues es el fundamento de toda la problemática sobre la supuesta viabilidad o inviabilidad de una inteligencia artificial general.

Asimov entendió y presentó la intuición como irreplicable, y como una manifestación inequívoca de la autenticidad humana. Tanto es así que en un momento de especial cansancio y dificultad para Baley durante el transcurso de la novela Robots del amanecer, el detective pronuncia estas palabras de tono afligido que, precisamente porque señalan la carga que supone ser un humano, subrayan muy bien estas diferencias:

Por primera vez en su vida, Baley sintió envidia de los robots. Se imaginó lo que sería poder caminar bajo aquel diluvio, ser indiferente al agua, a las imágenes, a los sonidos, ser capaz de hacer caso omiso de lo que le rodeaba y llevar una pseudovida

absolutamente valiente, y no conocer el miedo al dolor o a la muerte, pues el dolor y la muerte no existían. Y, no obstante, ser incapaz de poseer originalidad de pensamiento, ser incapaz de tener destellos imprescindibles de intuición… (RA, 339)

Hay numerosas partes constitutivas de la intuición a las que el robot no puede acceder para poder disfrutar de esta cualidad: las impresiones, los estereotipos, la habilidad para descartar las inferencias procedentes de lo que se conoce o se observa para adoptar nuevas conclusiones innovadoras. En buena medida, ninguna inteligencia artificial ha sido capaz de superar todavía un test de Turing completo precisamente por esta falta de la capacidad de intuir, es decir, debido a la ausencia de las inferencias de sentido común necesarias para articular un verdadero lenguaje natural.

Esta imposibilidad del robot para tener atisbos inspirados de solución recuerda al comentario pronunciado por el científico computacional Donald Knuth, recogido por Nick Bostrom en su ensayo Superinteligencia: «… que la IA haya tenido ya éxito haciendo básicamente todo lo que requiere “pensar”, pero haya fracasado en lo que la mayoría de la gente y los animales hacen “sin pensar” –eso, de algún modo, ¡es mucho más difícil!–» (22).

En un orden más formal, el científico informático Judea Pearl, laureado con los prestigiosos premios Turing, Rumelhart y Harvey Prize –los tres en 2011 por sus relevantes aportaciones en el campo de la inteligencia artificial– nos ofrece en su modelo de «la escalera de la causalidad» (23) los niveles de comprensión que pueden alcanzarse sobre la realidad: el nivel de asociación, que implica observar el entorno y establecer correlaciones entre los sucesos acaecidos; el nivel de intervención en el mundo, que implica actuar, es decir, conocer las causas de aquello que observamos para interactuar con propiedad; y un nivel superior de pensamiento contrafactual, que implica imaginar y entender con profundidad, al objeto de aprender de la experiencia para actuar de forma distinta o reformular nuevas comprensiones futuras.

Fruto de su labor continuada con sistemas de inteligencia artificial, Pearl se sitúa, como Erik J. Larson (24), en la posición de afirmar

que las máquinas que operan bajo la lógica de estos sistemas, tal y como están planteados ahora, todavía lo hacen en el escalón más bajo: establecen patrones de actuaciones en base a observaciones, buscan coincidencias y establecen relaciones y predicciones a partir de las bases de datos con las que se han entrenado. Bases de datos que conforman una representación del pasado, pero no del conjunto de una realidad que se expande a cada minuto. Y una realidad en última instancia impredecible.

Con todo, la inteligencia artificial lleva a cabo una operación como mucho inductiva, que poco tiene que ver con encontrar causalidades que permitan reaccionar ante sucesos inesperados –ante los cuales suele fracasar–, y mucho menos con proyectar o imaginar, con base en esas causas o transformaciones nuevos modos de actuación –innovar, improvisar, conjeturar–. «El aprendizaje automático solo puede ir a la zaga del flujo de nuestra experiencia, simulando (lo que esperamos que sean) regularidades útiles. Es la mente –no la máquina– la que marca el ritmo.» (25)

En un sentido similar, esta posición superior de la causalidad, del pensamiento contrafáctico y de la intuición por encima del mero cálculo sostenido en la observación de regularidades, vuelve a manifestarse en momentos narrativos en los que se baraja aplicar o no la justicia a rajatabla.

Para el robot Daneel –para más inri creado por los espaciales, una cultura entregada al más absoluto mecanicismo– solo hay una vía posible para solucionar los problemas, la justicia en su definición clásica: la permanente y constante voluntad de atribuirle a cada uno su derecho; dar a cada uno lo que le corresponde según sus actos, sin que medie ningún otro tipo de consideración. Esto es, atenerse a lo observado y registrado en su sistema, una operativa maquinal, cerrada y lógica; una operativa positiva, necesaria, pero incompleta según las circunstancias de ciertos casos.

Como se encarga de recordarle Baley a Daneel «hay algo superior a la justicia que se te ha incorporado. Existe un impulso humano conocido como piedad; un acto humano conocido como perdonar» (BA, 208) –por más señas, Asimov elige explicar estos conceptos, con los que la caridad supera a la justicia, sirviéndose del relato evangélico de la adúltera a la que Jesús salva de la lapidación, pues

Elijah Baley es creyente–.

Sin embargo, la piedad y el perdón no son «impulsos» que un robot sea capaz de entender por escapar al sistema de soluciones cerradas de su algoritmo; la respuesta de la máquina no puede ser otra: «No estoy familiarizado con esas palabras, compañero Elijah» (BA, 208).

«Al borde de lo incognoscible»

En la literatura de Asimov concerniente a los robots, las cosas que verdaderamente importan en la vida son accesibles, en exclusiva, para las personas. Una máxima aplicable no solo a las grandes aspiraciones y metas del hombre, sino también a los bienes relacionales cotidianos: una expresión improvisada de afecto, el abrazo de apoyo de un amigo, el beso cargado de deseo a quien se ama… Gestos y manifestaciones emocionales cuya expresión sincera corresponde tan solo al ser humano.

De hecho, en Las lunas de Júpiter, de la saga de Lucky Starr, uno de los métodos del protagonista para buscar a un posible robot espía que actúa como agente del mundo adversario de Sirio, en las filas de su expedición científica, es buscar a una persona sin emociones: «¿Acaso un hombre sin emociones no sería inmediatamente identificado como robot?» (LJ, 83).

Si acaso, poniendo algo de imaginación por nuestra parte y suspendiendo por unos momentos nuestra incredulidad, las emociones pueden ser imitadas por las máquinas a modo de pobre consolación. Y es que cualquier emulación sentimental de agrado por parte de un robot no es más que la fría ejecución de una ley programada de antemano: una Primera Ley que evita desagradar a través de la incomodidad y daño emocionales; una Segunda que impone la expresión de aprecio en cumplimiento de una orden – todo gracias a «pistas cerebrales positrónicas con orientaciones controladas por los potenciales y contrapotenciales establecidos por las Tres Leyes de la Robótica» (AS, 116); o la manifestación de «emociones sintéticas» previamente introducidas en su sistema mecánico por sus fabricantes con objeto de engañar a las personas. Sin embargo, aun con todo el empeño emulador, donde no hay voluntad ni deseo, los gestos pierden todo su valor:

[Baley] ordenó sus pensamientos y comprendió que no estaba abrazando a Daneel sino a R. Daneel, el robot Daneel Olivaw. Estaba abrazando a un robot y el robot le asía ligeramente, dejándose abrazar, estimando que la acción daba placer a un ser humano y tolerando esa acción porque los potenciales positrónicos de su cerebro le impedían rechazar el abrazo y causar de este modo una decepción al ser humano. (RA, 37)

De todo lo anterior se extrae una enseñanza sumamente valiosa para poner en su sitio cualquier exageración futurista sobre el imperio de los robots en nuestro mundo: que el robot, por muy evolucionado que se presente, nunca podrá superar al ser humano en todas sus facultades. El reducto emocional permanece impenetrable y marca, junto a otras características que estamos repasando en este capítulo, la diferencia entre el hombre y la máquina. Si son entes distintos, en último término no son comparables ni intercambiables en sentido absoluto.

A este respecto, destacan las palabras con las que Elijah intenta hacer ver a Francis Clousarr, el revolucionario «medievalista» que se posiciona contra los robots bajo la premisa última de que acabarán sustituyendo a las personas en todos los ámbitos, el absurdo de su temerosa suposición. Porque hay áreas de la comprensión y disfrute de la realidad que siempre permanecerán vedadas a la comprensión de una máquina. En uno de los diálogos más inspirados de su literatura, Asimov nos lo hace saber así a través de Elijah Baley:

Ni siquiera podemos construir un robot que sea tan bueno como un ser humano en todo lo que importa, y mucho menos mejor. No podemos crear un robot con sentido de la belleza, de la ética o de la religión. [...] No mientras existan cosas que la ciencia no pueda medir. ¿Qué es la belleza, o lo bueno, o el arte, o el amor, o Dios? Estamos constantemente al borde de lo incognoscible, e intentando entender lo que no puede ser entendido. Eso es lo que nos hace hombres. (BA, 221)

«Apreciar los problemas como un todo unido»

Dadas estas diferencias cualitativas de acceso a lo cognoscible, y de intuición o vivencia de lo incognoscible, al ser humano le queda reservada con permanencia la potestad de ofrecer algo único vedado a la máquina. Este planteamiento entronca a la perfección, desde su formulación a décadas vista, con las inquietudes presentes acerca de las supuestas facultades de la inteligencia artificial generativa para desterrar el gobierno de la inteligencia humana en ámbitos donde esta siempre ha sido el elemento central de sentido articulador de la actividad colectiva.

Nos referimos, sin duda, a la enseñanza y el rol que van a desempeñar estas herramientas tecnológicas en el aula y fuera de ella; en el trabajo de los estudiantes, pero también de los profesores. Los unos utilizan estas aplicaciones a las claras, sin esconderlo ni avergonzarse por ello; estudiantes de todos los niveles han adoptado estas herramientas e integrado su uso en el proceso de estudio, elaboración de trabajos y preparación de pruebas y exámenes con mayor rapidez que cualquier otro recurso formativo de búsqueda, creación y edición de la información.

La mayor parte de los otros, los profesores, se lanzan a criticar en masa esta adopción bajo la equivocada idea –de nuevo cunde aquí, incluso entre las mentes más pensantes, el mito de la inteligencia artificial– de que el soporte y las respuestas de estas herramientas están al nivel de la inteligencia humana –de la suya propia, de hecho–; de que pueden resolver todos los problemas con una pericia y grado de exhaustividad iguales a las de un investigador experto en la materia; y de que son capaces de responder con un alto grado de originalidad todas las preguntas. Lo que los profesores temen –una vez más el mismo miedo de siempre, tan humano– es que la inteligencia artificial les convierta en irrelevantes.

El pánico experimentado por esta supuesta amenaza puede ser percibido por ojos ajenos como dotado de todos los rasgos

caracterizadores de una profecía autocumplida: mientras la entrada de la robótica y de la inteligencia artificial en los centros educativos ofrecía una oportunidad de cambio y reacción para recuperar el verdadero espíritu crítico de carácter humanístico y las desaparecidas artes de la lectio y la disputatio que siempre bombearon el corazón mismo de la educación –competencia, actitud y artes en las que una inteligencia no humana tiene poco que decir o hacer–, los docentes malgastaban sus esfuerzos en combatir lo que no tenía que ser combatido sino reflexionado. Combatiéndolo se le dio la importancia que no tenía.

En cierta forma, se podría decir que esta irrupción de las inteligencias artificiales generativas en el debate educativo –en concreto de ChatGPT– fue contemplada hace más de ocho décadas por Asimov en su relato Corrector de galeradas (VR). La empresa U. S. Robots desembarca en el ámbito académico universitario con un corrector inteligente de textos, el robot EZ-27 (Easy). La finalidad de este proyecto es poner a prueba su robot con el público más exigente –prestigiosos docentes e investigadores– para la labor que debe realizar, la corrección de textos científicos; todo ello con objeto de validarlo y obtener un valioso refuerzo con vistas a su campaña comercial ante una opinión pública siempre escéptica.

Pese a las reticencias de la Universidad de Northeastern, el profesor Simon Ninheimer, una eminencia internacional en sociología acepta probar al robot en su labor de investigación. Sin embargo, termina denunciando a U. S. Robots bajo la acusación de que el robot ha reescrito uno de sus libros de acuerdo con las propias concepciones de la máquina acerca de lo que está bien o mal decir sobre otros.

La hipótesis de partida es que la Primera Ley habría obligado al robot a cambiar ciertas expresiones del texto original escrito por el profesor Ninheimer para evitar que los lectores pudieran sentirse dañados por las declaraciones del profesor –en una prefiguración de los «lectores de sensibilidad» que proliferan en la actualidad, y que ayudan a limpiar de posibles expresiones hirientes los libros antes de su publicación, con objeto de no dañar a ciertas minorías, en una expresión más de la cultura de la cancelación–.

La denuncia, que como demuestra la robopsicóloga Susan Calvin resulta ser falsa, pone sobre la mesa al menos dos temas del máximo

interés y dignos de ser reflexionados por su conexión con el momento presente.

El primero, el temor del profesor ante la precisión del robot, el cual es capaz de corregir los textos con mayor pericia no solo que su ayudante de investigación, sino que él mismo. Ninheimer ve en esta habilidad un riesgo para el devenir futuro de la intelectualidad universitaria y el sentido y supervivencia de la propia labor docente (VR, 451):

Sus robots se han apoderado de las galeradas. Muy pronto ellos, u otros robots, se apoderarán también de la escritura original, de la búsqueda de las fuentes, de comprobar y recomprobar los distintos pasajes, tal vez incluso de realizar las deducciones para las conclusiones. ¿Y qué le quedará entonces al erudito? Solo una cosa: las estériles decisiones relativas a las órdenes que habrá que dar al robot siguiente…

El segundo tema de interés para nuestros días es el temor, este sí, plausible, a que las corporaciones dominadas por intereses políticos, económicos, sociales o culturales de diversa índole, utilicen estos sistemas de generación de texto para introducir, ante la pasividad por defecto y la poca predisposición a esforzarse del usuario que da la orden de suministro de información, una cierta visión del mundo sobre lo que está bien, lo que está mal, lo que resulta correcto o incorrecto. No en vano, ChatGPT incluye entre sus advertencias de inicio el mensaje de que el contenido generado por su chatbot podría contener «instrucciones dañinas o contenido sesgado».

Sobre la primera de las preocupaciones, que advierte de la posible inutilidad parcial o total del profesor en el futuro dada la fastuosa capacidad de la máquina para generar, gestionar y transmitir información, Asimov vuelve en su breve ensayo Inteligencias unidas (VR (26)) a la distinción entre máquina y hombre señalada más arriba, centrada en la posesión de una «intuición», de una «penetración psicológica» –como el anhelo del saber y la curiosidad por conocer a toda costa–, o de la generación de nuevas

asociaciones conceptuales nunca disponibles.

Capacidades todas ellas imposibles para el robot o la inteligencia artificial, incluso en el mejor de sus desarrollos, aquel que tiene lugar fuera del ámbito de un algoritmo que lo constriñe, y que cuenta con el sistema de «aprendizaje profundo» –deep learning– y de millones de parámetros en cientos de capas para emular las conexiones sinápticas de nuestros cerebros.

Respecto al valor que los humanos realmente apreciamos en el término inteligencia, esta suele ser en concreto la capacidad de apreciar los problemas como un todo unido y homogéneo, la capacidad de captar las soluciones con la ayuda de la intuición y de la penetración psicológica, de apreciar la aparición de nuevos vínculos entre las cosas, el captar conceptos. ¿Seremos acaso capaces de programar un ordenador con estas características? Muy probablemente no. No tenemos ni la más mínima idea de cómo podría llegar a hacerse una cosa así. (VR, 519)

Décadas después de que Asimov escribiera esta reflexión, en un momento de pleno auge de la robótica y la inteligencia artificial, seguimos sin tener la más mínima idea de cómo podría llegar a hacerse una cosa así. Lo que sí sabemos es cómo no se logrará: persistiendo en el modelo actual de inteligencias artificiales basadas en inferencias lógicas. Quizá no solo no se pueda lograr, sino que incluso no haga falta que se logre nunca, si lo que hay sigue en proceso de mejora continua y ofrece prestaciones suficientes para servir de apoyo a la incansable tarea del progreso humano.

El lenguaje, que se construye en y de la experiencia, y no solo de las connotaciones y las acepciones registradas por un diccionario e indexadas por la máquina en su base de datos, adopta significados que son comprensibles en su totalidad –en el apartado cognitivo, pero también en el afectivo– solo por quienes disponen de ese contexto por haberlo vivido, terreno vedado a la máquina, que no «lo vive», sino que «lo indexa».

Un ejemplo aproximado de esta idea se encuentra en el distinto impacto que pueden llegar a causar en la percepción humana, más allá de posibles distinciones jurídicas, los términos «asesinato» y «homicidio». Una diferencia que el robot Daneel no logra captar:

—De las dos palabras, una de ellas helará la sangre de un ser humano más efectivamente que la otra, Daneel.

—¿A qué es debido?

—Connotaciones y asociaciones; el efecto sutil, no de la acepción del diccionario, sino de años de uso; la naturaleza de las frases, circunstancias y acontecimientos en los que uno ha experimentado el uso de una palabra en comparación con el de la otra.

—No hay nada de esto en mi programación —observó Daneel, con un curioso tono de impotencia sobre la aparente falta de emoción con que lo dijo (la misma falta de emoción con que lo decía todo). (RA, 41-42)

No habrá entonces una suplantación de una inteligencia por otra, sino una tendencia a posicionar cada una de ellas en la mejor versión de sí misma. Al menos en teoría, y si el rumbo de ambas responde a su propia naturaleza diferenciada, conforme la inteligencia artificial avance en sus competencias, los humanos dispondrán, gracias a ella, de mayores recursos para aprovechar e invertir en aquello que solo podemos hacer nosotros, aquello para lo cual la máquina se muestra incapaz, es decir, nuestra propia «capacidad de apreciar los problemas más globalmente aún» (VR, 520).

La forma en que Asimov especula y presenta las distintas maneras en que funciona la memoria en máquinas y personas en la novela Robots e Imperio puede ser ilustrativo de las posibilidades de complementarse de ambas inteligencias, cada una de ellas aportando lo mejor de sí misma.

Los robots, gracias a su atención teledirigida, a prueba de distracciones, y su gran capacidad de procesamiento y almacenamiento de la información en tiempo real, disponen de una memoria literal, en la que son capaces de recuperar todos y cada uno de los eventos materiales acaecidos en un momento y lugar con suma precisión, sin dejar escapar ningún detalle, pues todo lo registrado queda almacenado de manera indeleble y el paso del tiempo no borra ni diluye el recuerdo:

Los recuerdos que naturalmente no funcionaban según los sistemas humanos. Carecían de la rememoración imperfecta, de la impresión borrosa, de la adición y sustracción dictadas por anhelos y egoísmos, por no hablar de los deseos, lagunas y retrocesos que transforman el recuerdo en horas interminables de soñar despierto. (RI, 66-67)

Esta forma de memorizar puede ser un estupendo recurso para un registro y recuperación fidedignos y dotados de toda objetividad acerca de los acontecimientos pasados. En una memoria así no interferirían ni la natural interpretación de la persona que mira y que recuerda, ni sus olvidos –accidentales o intencionados–. Aprovechar esta «foto» del pasado con todo lujo de detalles y la imparcialidad de la que es capaz la máquina en su descripción de la misma puede ser un apoyo magnífico para la construcción del conocimiento, que necesita de una base sólida de datos e informaciones desde los cuales la humanidad pueda trabajar con garantías.

Este proceso es ejemplo de la especialización de cada una de las inteligencias, al mismo tiempo que una sirve de complemento a la otra.

Cada variedad de inteligencia tiene sus ventajas. La combinación de la humana con la de los ordenadores, coordina y elimina sus propias limitaciones, logrando una mejora general, gracias a la

unión de sus diferentes capacidades, mucho más rápidamente que intentándolo de manera individual. No será un caso de competencia entre una y otras negligencias, ni en absoluto una sustitución, será el complemento de ambas trabajando juntas más eficientemente que por separado, adecuándose siempre a las leyes de la Naturaleza. (VR, 520)

La expresión práctica más evidente de esta idea la obtenemos hoy en el arte generado por ordenador, con un amplio rango de aplicación que sobresale –está mucho más avanzado– en las artes gráficas y audiovisuales, donde la inteligencia artificial integrada en programas de creación y edición con una larga trayectoria, o la irrupción de nuevas aplicaciones de inteligencia artificial generativa hacen de la máquina una herramienta que supera toda equiparación con las que venían utilizándose para estas actividades artísticas.

En estos casos, la máquina no equivale a un cincel, una cámara fotográfica o de vídeo, un pincel, un lápiz, un rotulador o incluso un software de edición al uso. La máquina es todo eso, en el sentido de que es una herramienta, y algo más que eso, dado que introduce un refuerzo de la capacidad de creación y edición del artista.

Capacidades adicionales y diferentes que ofrecen como resultado una obra final que ya no es fruto exclusivo de las facultades del artista, pues la obra depende, aunque fuera en grado menor, pero depende en algún punto a nivel creativo de la herramienta.

Este es un ejemplo de complementariedad de ambas inteligencias, gracias al cual las personas podemos encontrar nuevos derroteros desde los que expandir nuestra creatividad o, en un sentido menos directo y elevado, reducir gracias a la automatización los tiempos de trabajo de las tareas de edición rutinarias en las que no se pone en juego ninguna cualidad artística superior.

«Pensar en las preguntas adecuadas»

En todos los procesos mentados con anterioridad la inteligencia artificial no menoscaba ninguna facultad esencialmente humana. Las personas siguen desempeñando un papel guía fundamental en la configuración del mundo, pues conservan para ellas la misión imprescindible de otorgar un propósito personal o colectivo a aquello que se emprende en respuesta a los interrogantes que nos conmueven.

Puede tomarse en este sentido esta célebre e inspiradora cita del relato Se puede evitar el conflicto: «He observado que los hombres capacitados siguen estando solicitados en nuestra sociedad; todavía necesitamos al hombre que es suficientemente inteligente como para pensar en las preguntas adecuadas» (VR, 245).

También la constante actitud interrogativa de la pareja protagonista de Los límites de la Fundación y Fundación y Tierra, Golan Trevize y Janov Pelorat, quienes trazan su búsqueda de un planeta Tierra olvidado y perdido en los confines cada vez más inexplorados de la Galaxia, todo a base de preguntas: «Lo que cuenta es que he descubierto la pregunta que debo hacer y, naturalmente, una buena pregunta es el medio para obtener infinitas respuestas», dice Pelorat (LF, 321).

El deseo de una mayor comprensión de la realidad, el querer saber cuya única posibilidad de formulación reside en preguntarse y preguntar, en la autoexigencia nacida de lo más profundo, no es facultad de las máquinas, y es la potencia que nos mueve como individuos y como sociedad.

En refuerzo de esta idea, debemos recordar que una de las grandes y más icónicas protagonistas de la narrativa de Asimov dedicada a los robots, la primera gran profesional de la robopsicología de la historia, Susan Calvin, debe toda su fama y buen hacer en la resolución de los difíciles casos que se le plantean a su prodigiosa

habilidad para formular esas preguntas adecuadas. Susan Calvin es graduada en Cibernética, pero –detalle no menor para el caso que queremos subrayar– se doctoró en Filosofía.

La cuestión llevada a su sentido más práctico, pues también atañe en gran medida a la previsión y supervivencia de ciertas profesiones en el futuro, aparece con frecuencia en la literatura de Asimov revisada en este libro. En el relato Algún día, el joven Nickie quiere ser programador cuando sea mayor, y cuenta que el señor Daugherty, el maestro de la escuela, le ha dado el siguiente consejo: «Cualquiera puede supervisar en los controles, comprobar las respuestas y resolver problemas de rutina. [...] el truco está en ampliar la investigación y encontrar formas de hacer las preguntas adecuadas, y esto es difícil» (VR, 339).

En este fragmento vuelve a ponerse de relieve la idea de que es nuestra capacidad para interrogarnos, interrogar a otros y cuestionar la realidad lo que posibilita ofrecer sentido al presente, aquello que logra diferenciarnos de la inteligencia artificial, pues a las preguntas siguen las respuestas y los propósitos para aquello que se emprende y que otorgan un sentido a nuestra vida.

Y a la mención del propósito cabe sumar el recordatorio de que, aunque la inteligencia artificial esté o estuviera capacitada para igualar o superar el desempeño humano en la ciencia o el arte, esta superioridad se manifiesta en áreas muy específicas y tareas muy concretas. La capacidad humana para la generalización, y la presentación y tratamiento de los problemas con una visión global queda intacta.

Ambas cualidades, plantear las mejores preguntas y pensar de manera transversal y holística, obtienen cada vez más importancia, tanto más cuanto más avanza la inteligencia artificial. Cuanto más peso adquieren las cualidades de esta, más relevancia adquieren aquellas.

Esta situación revaloriza una vez más el papel que desempeñan las humanidades –con las múltiples disciplinas que la integran–, que en su ofrecimiento de las mejores cualidades de la inteligencia humana queda blindada frente a la inteligencia artificial. Los trabajos con un componente fuerte de las humanidades son los más seguros. Sus

funciones nunca podrán ser desplazadas por las máquinas ni quedarán obsoletas frente a los avances tecnológicos.

No caben, por tanto, los lamentos por una sustitución de las capacidades esenciales de la inteligencia humana. Más bien la inteligencia artificial subraya y nos ayuda a valorar lo propia y exclusivamente humano que sobresale frente a aquellas competencias puramente maquinales que sí pueden ser ejercidas por el robot u otras máquinas y para las que, en definitiva, no se necesita al hombre.

Aunque en la saga de la Fundación solo aparecen los robots hacia el final de la misma, y el número y profundidad de las reflexiones acerca del tema en esos volúmenes no aporta demasiada novedad con respecto al resto de la obra de Asimov dedicada a la robótica, no se puede pasar por alto que el valor de la inteligencia humana, como principio y poder superior a cualesquiera otros de carácter tecnológico y material, es una de las ideas más importantes que se desprenden de la lectura del conjunto de su serie literaria más célebre, idea que toma más y más fuerza conforme avanza el desarrollo de la misma.

No tanto el acopio de información y la recopilación del conocimiento, simbolizados en la Primera Fundación a través de la gigantesca tarea de confeccionar una Enciclopedia galáctica –tarea que pronto se descubrirá como un tremendo macguffin de Hari Seldon–, como el ejercicio intelectual, el poder de la mente y de la sabiduría como instrumentos rectores del rumbo de la humanidad, tanto para bien –como parece que pretenden los sabios de la Segunda Fundación y los agentes de Gaia– como para mal –el célebre personaje de El Mulo– (27).

Idea que también podemos encontrar expresada de manera explícita en esta misma saga, cuando la supuesta hameniana Sura Novi, campesina habitante de Trántor, se sorprende al descubrir que los maestros de la Segunda Fundación que habitan su planeta, considerados magos cuando no cuasi dioses, disponen, sin embargo, de una tecnología material muy pobre en comparación con la Primera Fundación. En una conversación con el maestro Gendibal acerca de la imposibilidad de dar caza a la nave gravítica de la Primera Fundación de Golan Trevize, Gendibal valora lo

importante:

—¿Más rápida que una nave de los maestros? —Sura Novi pareció consternada. Gendibal se encogió de hombros.

—Como tú dices, es posible que yo sea un maestro, pero no lo soy en todo. Los sabios no tenemos naves como estas, ni tenemos muchos de los dispositivos materiales que poseen los dueños de esas naves.

—Pero ¿cómo pueden los sabios carecer de tales cosas, maestro?

—Porque somos maestros en lo que es importante. Los progresos materiales que tienen estos otros son bagatelas. (LF, 385-386)

En otras ocasiones podemos encontrar expresada esta idea marco con rotundidad y desenfado, como en esta máxima de la Segunda Fundación: «El necio más incorregible es aquel que desconoce su sabiduría» (28).

▪

19 Boden, op. cit., p. 11.

20 Bostrom, op. cit., pp. 185-208.

21 Una habilidad a la que Asimov concede especial importancia, si tenemos en cuenta que también la posee en grado sumo Golan Trevize, último gran héroe de la saga de la Fundación –en los volúmenes de la bilogía final, Los límites de la Fundación y Fundación y Tierra–. Dirigiéndose a su compañero de misión, el historiador Janov Pelorat: «Janov —dijo—, tengo una corazonada al respecto. Poseo este don de llegar a una conclusión correcta

partiendo de casi nada. Siento una especie de seguridad en mi interior que me dice cuándo tengo razón… y ahora estoy seguro» (LF, 354).

El lector descubrirá que es bajo esta cualidad de la percepción que se desencadenan los acontecimientos de mayor repercusión en la trama y hasta para la Galaxia representada en la ficción.

22 Bostrom, op. cit., p. XX.

23 Pearl, y Mackenzie, El libro del porqué: la nueva ciencia de la causa y el efecto.

24 Larson, op. cit., pp. 157-158.

25 Ibid., p. 169.

26 De la serie de reflexiones de corta extensión con la que finaliza el conjunto de relatos de esta edición.

27 La valía y el poder del control mental están igualmente presentes en varias de las aventuras de la saga de Lucky Starr, como en la captación del pensamiento de los marcianos que el protagonista descubre en El ránger del espacio en el subsuelo de Marte, o el control mental de las V-ranas de Los océanos de Venus, criaturas cuya facultad queda cerca de causar el colapso de toda una civilización.

28 Asimov, Trilogía Fundación, p. 681.

Los riesgos del progreso robótico

De la tesis defendida por Asimov, según la cual robots y humanos, inteligencia humana e inteligencia artificial son esencialmentedistintos y, lejos de entrar en competencia pueden combinarse en su mejor versión para alcanzar nuevas cotas de progreso, no se sigue que solo por existir esta posibilidad el ser humano vaya a aprovechar todo su potencial.

En la literatura de Asimov podemos tener un atisbo de planetas y sociedades futuristas. En la saga de los robots podremos visitar Solaria y Aurora como muestra de los Cincuenta Mundos Exteriores, la civilización humana que dio el paso a conquistar el espacio sirviéndose del potencial de los robots. Podemos atestiguar el rumbo poco ilusionante y prometedor que ha tomado el desarrollo de las sociedades espaciales, instaladas en un individualismo extremo y egotista, y configuradas en una actitud pasiva hacia los problemas de terceros, bajo la máxima de que la tecnología provee para todos.

En Solaria, el mundo descrito en la novela El sol desnudo, la robótica ha alcanzado el mayor nivel técnico de toda la galaxia y puede permitirse vivir exclusivamente de la venta de sus máquinas positrónicas. En Solaria hay más robots que humanos y estos sienten un fuerte rechazo ante la posibilidad de tener contacto con sus congéneres. ¿Para qué encontrarse en persona si se dispone del «triménsico», una tecnología generadora de realidad aumentada que duplica y reproduce con toda fidelidad a ambos sujetos a distancia como si el encuentro fuera real?

Los matrimonios son concertados trazando la viabilidad y calidad genética de la descendencia. ¿Para qué permitir relaciones libres, basadas en el amor y sostenidas en el compromiso, si se puede prever de antemano con precisión de laboratorio el «resultado» de una unión, logrando además un buen índice de «limpieza genética»?

La población de Solaria está firmemente limitada a un máximo inamovible de veinte mil personas. ¿Para qué más, si este es el número de ciudadanos necesarios para sostener en equilibrio matemático el avance científico y estético de la robótica que permite dinamizar la economía del planeta?

Los solarianos apenas salen de los muros de sus «establecimientos» para transitar un paisaje exterior por lo demás desolado. ¿Para qué deberían salir de sus casas si la robótica les provee de todo lo que necesitan? Y los embriones se incuban en máquinas. ¿Para qué dejarlos en el seno materno si una máquina de laboratorio puede hacerlo sin molestias para los padres y en caso de desperfectos genéticos o complicaciones durante la gestación los científicos pueden elegir eliminarlos con todas las facilidades?

La situación en la que ha terminado estancada Solaria no es muy diferente a la de Sirio, confederación adversaria de la Federación Terrestre en la saga de Lucky Starr. La Confederación Siriana es un reflejo perfecto de Solaria en los aspectos sociológicos que acabamos de mencionar, lo que, contando con que ambos mundos pertenecen a series literarias distintas, reforzaría la observación de que Asimov pareciera querer alertarnos de los posibles riesgos de estar excesivamente preocupados y centrados en su desarrollo tecnológico. Así lo expone el doctor Peverale en El gran sol de Mercurio: «Los hombres que se acostumbran a preocuparse por las necesidades de unas máquinas, se vuelven insensibles respecto a las necesidades de los hombres» (GSM, 61).

Los planetas de Sirio están escasamente poblados y extremadamente descentralizados. Viven en unidades aisladas de familias individuales, cada una de ellas con su propia fuente de energía y servicios. Cada una tiene su grupo de esclavos mecánicos no hay otra palabra posible, esclavos en forma de robots positrónicos, que hacen el trabajo.

[...]

Y, de todos modos, la vida humana, cualquier clase de vida humana, no significa demasiado para ellos. Su civilización está

centrada en las máquinas. Para ellos tiene más importancia un robot siriano que un hombre siriano. Consideran que un robot vale tanto como cien hombres de la Tierra. Miman a esos robots. Los quieren. (GSM, 58, 61)

Es más, la situación en la que ha terminado estancada Solaria es el resultado de circunstancias y condiciones muy parecidas a las descritas hoy por los teóricos de la singularidad –el matemático Vernor Vinge y el ingeniero Raymond Kurzweil, entre los más destacados representantes de esta corriente–: el hipotético punto de no retorno en el que la inteligencia artificial fuerte alcanzará niveles de desarrollo y potencia por encima de la inteligencia humana, convirtiéndose así en «inteligencia artificial sobrehumana», es decir, el momento en el que las máquinas serán más inteligentes, y por tanto superiores al ser humano, conduciendo a una posible hibridación máquina-hombre o a la superación de la especie humana –o incluso la superación y sustitución de la realidad física por una realidad virtual–, ideas estas últimas también recogidas en la filosofía transhumanista –filosofía, por lo demás, de marcado carácter futurista: dado que estamos muy lejos de lograr una inteligencia artificial fuerte, sobre otra de carácter sobrehumano solo cabe especular–.

Sin entrar a fondo en la cuestión, la singularidad conduce a un mundo en el que la evolución tecnológica ha superado a la evolución biológica (Boden, 2017), precisamente una situación con la que Solaria, como presentábamos más arriba, guarda grandes similitudes.

En este mundo de ficción creado por Asimov, todo se ha confiado a la robótica, perdiendo de vista, eso sí, que la robótica solo es capaz de ejercer y representar la parte más instrumental de las sociedades humanas, pero no es capaz de aproximarse, ni de lejos, a la verdadera condición humana, que nada tiene que ver con el rendimiento o la maximización de resultados aplicada a la vida y a las emociones de las personas.

Lo que en definitiva parece querer ilustrar Asimov con la proyección a futuro de la humanidad en estas condiciones de honda

y acelerada deshumanización es que, si caemos en manos de los robots, no será tanto porque las máquinas nos ataquen y suplanten, como porque nos abandonemos a ellas en demasía, perdiendo nuestros deseos, motivaciones y objetivos –convirtiéndonos nosotros mismos, en el sentido anímico del término, en máquinas–.

No en vano, una novela sumamente inteligente como Sinsonte (1980) de Walter Tevis, capaz de aglutinar todas las grandes propuestas distópicas del siglo XX, basó su defenestrada civilización futura, a punto de la extinción, en la idea de que la creciente dejadez y pasividad humanas habrían llevado a las personas a abandonar cualquier propósito vital, y a depositar cada vez más responsabilidades en los robots hasta llegar a un punto de no retorno en el que ni siquiera leer y escribir era posible. La alienación no viene de una dictadura o régimen totalitario instaurado por los robots o la inteligencia artificial –en esta historia los robots solo mantienen el mundo que se les ha legado sin que lo pidieran–, sino de la indolencia de una sociedad que ha soltado las riendas.

Personas robotizadas

La consecuencia más inmediata de esta dejación de funciones, este poner en manos de los robots realidades que una máquina es incapaz de comprender y manejar, es la creación de un mundo en el que las personas han sido «robotizadas» y actúan, a todos los efectos, como lo haría una máquina; o bien solo actúan las máquinas, pues los seres humanos han quedado inutilizados, no son capaces de concebir un mundo sin robots en el que no sean estos no solo los que trabajan, sino los que velan por el orden moral. Esto es, un mundo deshumanizado en el que, por muchas comodidades de que se dispongan frente a un planeta Tierra atestado, atrasado y no robotizado, no merece la pena vivir.

Así cobran sentido las terribles palabras del doctor Levular Mandamus en la novela Robots e Imperio al valorar la existencia en ambas situaciones: «La gente de vida breve de un mundo enfermo y superpoblado estaba mucho mejor muerta... Cien veces mejor muerta que viviendo la vida que ellos se habían obligado a vivir» (RI, 275).

La misma lógica expresada por el doctor Mandamus impera milenios más tarde, cuando volvemos a visitar Solaria en la saga de la Fundación, en su última novela, Fundación y Tierra, encargada de unir esta serie literaria con el ciclo de los robots. Allí encontramos a Sarton Bander, humano hermafrodita. En el lapsus temporal de esos milenios transcurridos, los solarianos han trabajado en aunar los cuerpos de hombre y mujer mediante el control genético robotizado y han eliminado el sexo biológico, acabando así con la práctica totalidad de los rasgos identitarios humanos. Esgrimiendo un concepto extremadamente libertario de la libertad, Bander proclama:

Siempre habéis vivido en enjambres [en referencia a la raza

humana «tradicional»] y no conocéis otro estilo de vida que el de sentiros obligados constantemente, incluso en las cosas más pequeñas, a doblegar vuestra voluntad a la de otros, o lo que es igualmente vil, a pasaros la vida luchando por doblegar la voluntad de los otros a la vuestra. ¿Es eso libertad? ¡La libertad deja de serlo si uno no puede vivir como quiera! ¡Exactamente como quiera! (FT, 261)

Tampoco entendéis la libertad. Si existe otro solariano que pueda disputarme mi absoluto dominio sobre cualquier trozo de mi tierra, sobre cualquier robot o cosa viviente u objeto, mi libertad queda limitada. Y como existen otros solarianos, la limitación de la libertad debe ser eliminada todo lo posible separándoles hasta el punto de que el contacto sea virtualmente inexistente. (FT, 263)

En las visitas narrativas a Solaria de la mano de Asimov comprendemos que la libertad entendida como el ejercicio de la voluntad sin límites ha conducido al borrado del otro, al borrado de uno mismo –la inexistencia de una identidad propia, siquiera biológica, natural– y a una existencia solitaria y aislada del mundo donde todo está organizado y ordenado al milímetro, sin espacio para la improvisación, la causalidad ni el asombro para conseguir el máximo de eficacia y efectividad en el logro de un placer constante, ininterrumpido para uno mismo. Una existencia a todas luces maquinal, robótica.

Volviendo a la saga de los robots y sus personajes, cuando el detective terrícola Elijah Baley tome contacto directo con la sociedad espacial y se sorprenda frente a un modo de vida inconcebible para alguien que todavía conecta con la verdadera experiencia de la humanidad, el sociólogo solariano Anselmo Quemot, habituado a esta manera autolimitada de vivir, normalizará la situación denominando el estado de aislamiento y ataraxia perpetuo en el cual han quedado sumidos los habitantes de Solaria como «equilibrio social»: la pertenencia de todos los seres humanos a una clase social privilegiada donde la búsqueda de la vida, la libertad y la felicidad habrán tocado fondo porque los hombres tendrán todo cuanto deseen y necesiten gracias al trabajo de los robots, y habrá llegado así el fin de la historia humana (SD,

159-160).

Las palabras del sociólogo Quemot no son pronunciadas en tono de queja o lamento. Inmerso en una permanente dinámica de despreocupación y de disfrute de los placeres proporcionados por un mundo plenamente robotizado, sus palabras son la constatación de una realidad, al mismo tiempo que una advertencia para todas las sociedades de la Galaxia: el fin de la historia, la ausencia de propósito y el sinsentido de vivir sin una misión en la vida es el destino ineludible que espera a la humanidad cuando esta malinterpreta la evolución de la robótica como una oportunidad de desembarazarse de su naturaleza y dejar en manos de las máquinas aquello que en realidad solo puede acontecer en el corazón humano. Una humanidad en este estado deja de interrogarse, de aspirar a dar respuesta a las grandes preguntas y a sus más añorados anhelos.

Para terminar de cerrar el círculo, estas palabras son también proféticas y significativas en el conjunto de la saga literaria de los robots, en la que se enmarca El sol desnudo. Lo son en la propia novela, porque un asesinato viene a perturbar esa supuesta paz perpetua, recordando a todos los personajes por la vía de la violencia que la vida humana no está sujeta por mandato a mecánicas plenamente controlables y predecibles –asesinato para cuya resolución necesitan de un terrícola, el habitante de un mundo todavía meridianamente despierto–.

Lo son también en el transcurso de la serie literaria, dos novelas después, en Robots e Imperio, que transcurre dos siglos más tarde de los eventos narrados en El sol desnudo, porque comprobaremos que Solaria ha sido el primer mundo en extinguirse. E incluso proféticas para lo que vendrá años más tarde cuando Asimov retome la saga de la Fundación, y ambas series literarias, robots y Fundación terminen de unirse en la novela Fundación y Tierra.

En este relato el doctor Deniador, aunque sin pruebas, porque la memoria de aquel mundo ha acabado por desaparecer y borrarse en los tiempos de la Primera Fundación, lanza la suposición más acertada: «Es fácil deducir que una sociedad que depende por completo de los robots se vuelve muelle y decadente, debilitándose y muriendo de puro aburrimiento o, más sutilmente, por perder la

voluntad de vivir» (FT, 154).

En efecto, volviendo a la saga de los robots y El sol desnudo, el sociólogo Quemot no yerra en su descripción de Solaria. Dejar todo en manos de los robots puede suscitar el ideal del placer y el goce constante ante la ausencia de trabajo y de esfuerzo. Pero esa ausencia es en realidad una condena para los hombres, que necesitan por su propia configuración y condición de esas dificultades para dar aliento a su vida. En ausencia de las cuales, solo les espera su propia destrucción. Una humanidad así pierde lo más importante, como se encarga de resumir el detective Baley al subsecretario Albert Minnim a su regreso a la Tierra tras su paso por Solaria:

—Los solarianos han renunciado a algo que la humanidad ha poseído durante un millón de años; algo de más valor que la energía atómica, las ciudades, la agricultura, las herramientas, el fuego y que todo, porque es algo que hizo posible todo lo demás.

—No me gusta jugar a las adivinanzas, Baley. ¿Qué es?

—La tribu, señor. La cooperación entre los individuos. Solaria ha renunciado por completo a ella. Es un mundo de individuos aislados y el único sociólogo del planeta se halla encantado de que así sea. [...] La influencia humana ha desaparecido. El único futuro que prevén es la ectogénesis y el aislamiento completo desde la cuna.

—Todo eso es horrible —dijo Minnim—. Pero ¿usted lo considera perjudicial?

—En mi opinión, sí. Sin la interdependencia humana, desaparece el principal aliciente que ofrece la vida; se esfuman casi todos los valores intelectuales y falta una auténtica razón para vivir. La visualización no puede sustituir la presencia personal. Incluso los propios solarianos se dan cuenta de que la visualización no es más que un sentido a larga distancia.

Y si el estar aislados no bastase para producir el anquilosamiento, tendríamos la cuestión de su longevidad. En la Tierra, tenemos una

aportación constante de vidas jóvenes que introducen los cambios, porque aún no han tenido tiempo de anquilosarse. Supongo que debe existir una edad óptima: una vida lo bastante larga para realizarse plenamente y, sin embargo, lo bastante corta para dejar paso a los jóvenes a un ritmo que no resulte demasiado lento. En Solaria, ese ritmo es lentísimo. (ESD, 271)

Intimidad artificial

Otra evidencia del alcance de esta forma hiperindividualizada de vivir que describíamos más arriba, en la que toda experiencia acaba siendo mediada por una máquina, es su fuerte y negativo impacto en el factor relacional. La máquina siempre ofrece aquello que pedimos, cumple con todos nuestros deseos, en apariencia nunca decepciona porque nos proporciona todo aquello que queremos en el momento y cantidad en que lo queremos. En la relación con la máquina no hay riesgo alguno.

Por ello, al menos a priori, la relación con una máquina puede parecer, y de hecho ser –aunque a un nivel muy elemental–, mucho más satisfactoria que cualquier relación que se pueda mantener con otras personas, siempre susceptibles de negar su colaboración en las aspiraciones y proyectos del resto. Sin embargo, ¿en qué condiciones busca y otorga la máquina el bien para el otro? La máquina opera en régimen de obligatoriedad, está programada para cumplir. No conoce la libertad y, por ende, tampoco la gratuidad, y donde no hay gratuidad no puede haber una relación constructiva, basada en la verdadera amistad, mucho menos en el amor.

La preferencia por la máquina en todas las dimensiones de la vida, con especial hincapié en las relaciones afectivas, presenta su mejor ejemplo en el personaje de Gladia Delmarre. Como solariana, Gladia ha sido educada en el rechazo al contacto físico con sus semejantes y en la adopción de una actitud reservada frente a otras personas, incluso si los encuentros se realizan a través del triménsico. Por contraste, tras el asesinato de su marido migra y se establece en Aurora y descubre en este mundo una fuerte tendencia al libertinaje sexual. En Aurora, el sexo se puede tener con cualquiera, aunque se esté casado, y «casado» es convivir tres meses; el incesto, aun no siendo habitual, no está mal considerado.

Frente a este profundo choque cultural con respecto a su mundo de origen en materia sexual, prevalece, sin embargo, el talante distante

y frío entre personas, resultado del férreo individualismo generalizado en todas las sociedades espaciales, una característica común de ambos mundos que el sexo libre aurorano no parece remediar y el matrimonio aurorano tampoco, dado que si bien está abierto a la vida –siempre a través de fecundación artificial–, los lazos familiares no se consideran relevantes y muy pocos habitantes de este mundo conocen a su familia inmediata.

La enorme decepción de Gladia, que experimenta un fracaso tras otro en su intento de establecer algún tipo de relación verdaderamente humana, la lleva a buscar consuelo en el robot humaniforme Jander Panell, con el que mantendrá una relación afectivo-sexual e incluso contraerá matrimonio.

La historia recuerda, a la inversa –por la trasposición de los sexos entre el ser nacido y el ser creado–, al mito de Pigmalión, el escultor que talló a la mujer ideal con un objetivo sexual, y con la aspiración de que este ser artificial llegara a ser más perfecto que cualquier mujer de verdad. Pero cuenta más como trasunto de la tragedia de Laodamia, cuyo amado marido, Protesilao, murió en la legendaria Guerra de Troya. Atrapada en el recuerdo y la pena por la pérdida, Laodamia manda esculpir una estatua de su marido en bronce para entregarse a ella (29).

En Los robots del amanecer, la agalmatofilia de la Antigüedad representada por Pigmalión y Laodamia deriva en la robofilia de Gladia, quien también busca consuelo afectivo tras el asesinato de su marido y el frío trato y educación recibidos en Solaria. Gladia cree encontrar en un robot todo aquello que los hombres espaciales no son capaces de proporcionarle: un compañero de vida comprometido con su bienestar, un otro en el que encontrar afecto y comprensión, y una forma de satisfacer su deseo sexual.

Sin embargo, hay un «pero» en esta sustitución del hombre por la máquina, un pero reconocido por la propia Gladia al referirse a sus vivencias con Jander: al tratarse de un robot, todo aquello que hacía por ella tenía una sola razón de ser, la obligación y el deber impuestos por las leyes de la robótica; «no dar placer cuando podía hacerlo habría sido desilusionar. La desilusión podía ser considerada como un daño y él no podía dañar a un ser humano» (RA, 155). Una vez desaparecido Jander, cuya «muerte» o

«roboticidio» centra la trama e investigación de Elijah Baley en Los robots del amanecer, Gladia juzgará así su relación con el robot ante las preguntas del detective (RA, 155-156):

—Así pues, ¿fuiste completamente feliz?

—¿Con Jander? Por supuesto. Completamente.

—¿Nunca os peleasteis?

—¿Con Jander? ¿Acaso habría sido posible? Su única meta, la única razón de su existencia, era complacerme.

—¿No te sentías molesta por ello? Solo te complacía porque tenía que hacerlo.

—¿Qué motivo tenemos para hacer algo más que, por una u otra razón, tener que hacerlo?

—¿Y nunca experimentaste la necesidad de intentarlo de veras… de intentarlo con los auroranos después de haber aprendido a tener un orgasmo?

—Habría sido un sustituto insatisfactorio. Yo solo quería a Jander… ¿Entiendes ahora lo que he perdido?

De las palabras de Gladia podemos observar que el «matrimonio» que ha mantenido con Jander está muy alejado de ser tal. Más bien se trata de una relación unilateral en la que una máquina sirve a su dueña como ella quiere ser servida, y cuyo equivalente, en caso de tratarse de una relación entre personas, estaría muy cerca de considerarse la peor de las servidumbres o, directamente, esclavitud. Para Gladia todo se justifica en la visión de que solo hay una motivación posible para todas las acciones emprendidas, incluso por las personas: el «tener que hacer», el deber ser. Bajo esta forma de ver la vida, no le falta razón al considerar equivalentes las relaciones mantenidas con personas y con robots. Al descartar la solidaridad o la gratuidad como factor determinante de las mismas,

se suprime cualquier valor diferencial humano, y todo queda reducido a una sola característica, la obligatoriedad de un encargo que bien podría ser desplegada por un robot como por una persona, tanto más da.

El paralelismo más inmediato que a muchos aficionados al cine les vendrá a la cabeza al considerar el caso de Gladia es la película Her (Spike Jonze, 2013), en la que su protagonista, Theodore Twombly (Joaquin Phoenix) mantiene una relación sentimental con una inteligencia artificial tras su divorcio y la consiguiente fragilidad y soledad que experimenta.

Es un magnífico ejemplo, sin duda, aunque podríamos pensar que entonces todo queda en la literatura o el cine, es decir, en la ficción, cuando es precisamente en esta década de los veinte del siglo XXI cuando estamos experimentando el nacimiento y primer éxito de software dotado de inteligencia artificial para tener este tipo de relaciones. Aplicaciones como Replika, Blush, Xiaolce o CarynAI aglutinan millones de usuarios a los que, a través del análisis algorítmico de sus datos generados en el tiempo, son capaces de «conocer» y permiten, según el nivel de la suscripción contratada, mantener conversaciones a distintos niveles de intimidad para llegar a tener un novio o novia virtual –y artificial, claro–. Están diseñadas, como cualquier otra app del mercado, para satisfacer los requerimientos y preferencias del usuario; pero en este caso, esas prioridades conectan con las emociones más profundas de la persona, creando entre persona y máquina una intimidad que es, nunca mejor dicho, artificial.

En la literatura de Asimov encontramos numerosos relatos en los que se replica este simulacro de relaciones en el ámbito de los cuidados. La mayor parte de estas narraciones plantea la actuación de un robot como niñera que acaba por hacerse «amigo» del niño al que supervisa. Un argumento todavía reservado en toda su extensión para la ficción, pero que empieza a vislumbrarse como una preferencia de los padres cuando intentan aliviar la carga en el cuidado de sus hijos más pequeños y dependientes con la ayuda de un robot. El tema también cobra forma, y fuerza, al pensar en el cuidado de los mayores, donde el uso de robots para tareas asistenciales y el acompañamiento sí empieza a ser notorio. Para

todos estos casos, relaciones de índole sentimental o sexual, cuidados a menores y mayores, aplica una misma conclusión: son relaciones adulteradas, donde no hay verdadero cariño, aprecio, compañía o amistad, no digamos amor.

Es un engaño sumamente pernicioso, pues como plantearon hace años los investigadores Robert y Linda Sparrow en una reflexión imperecedera, «lo que la mayoría de nosotros queremos de la vida es ser amados y cuidados, y tener amigos y compañeros, no simplemente creer que somos amados y cuidados, y creer que tenemos amigos y compañeros, cuando en realidad estas creencias son falsas. Es decir, deseamos que el mundo real sea de cierta manera y no solo que nuestras creencias o experiencias sobre el mundo sean de cierta manera» (30).

Ambos autores terminan señalando que, para vivir queriendo que las cosas sean, y no solo que parezca que son, a nadie le gustaría cambiar su vida en el mundo real por una vida en un simulador de realidad virtual. Una idea muy presente en el conocido ensayo Anarquía, Estado y utopía de Robert Nozick, clásico contemporáneo del pensamiento político en el que, para plantear la problemática del contrato social y la búsqueda de la felicidad se nos presenta la metáfora de las «máquinas de experiencias». Un tanque en el que el sujeto flotaría conectado a electrodos que le suministrarían una experiencia réplica de la realidad vívida y fidedigna –máquina de experiencias– donde ser el tipo de persona que se elija ser –máquina transformadora– y donde conseguir todo aquello que queremos –máquina de resultados–.

¿Por qué no habríamos de ser felices en un entorno así, en el que parece que disponemos de todo lo que queremos y parece que podemos ser quienes aspiramos a ser? Precisamente porque lo que «parece» no «es». Obnubilados por la fastuosa promesa de la máquina, pasamos por alto que el hombre quiere hacer cosas, y no solo tener la experiencia de hacer cosas, que queremos ser cierta persona, y no solo tener la experiencia de ser cierta persona, y que queremos conseguir por nosotros mismos esas cosas, y no solo tener la experiencia de que conseguimos esas cosas.

En definitiva, como sentencia Nozick, «lo que más perturba de ellas [de estas máquinas] es que vivan nuestras vidas por nosotros. [...]

Quizás lo que deseamos es vivir (verbo activo) nosotros, en contacto con la realidad (y esto, las máquinas no pueden hacerlo por nosotros)» (31).

La máquina de experiencias, aun cuando es una metáfora con propósito explicativo, es también el perfecto equivalente al ejercicio indefinido de la suspensión de la incredulidad al que se prestan quienes aceptan o validan la idea de que los cuidados suministrados por un robot, o las relaciones sentimentales mantenidas con una inteligencia artificial avanzada pueden tener valor de equivalencia con los cuidados o las relaciones con nuestros iguales, es decir, con los cuidados o las relaciones de verdad. La aceptación y normalización de una mentira así nos mantiene atrapados en una ficción, inmensos en un simulacro permanente de falsos cuidados y relaciones, apartados y ciegos frente a la realidad; una situación no muy distinta a vivir una realidad paralela en ese tanque y con esos electrodos.

El cuidado que puede proporcionar un robot, como cualquier respuesta de una máquina a una demanda emocional, es y será siempre incompleto e insuficiente, dado que incluso cuando se produjera una simulación en extremo cuidada en sus detalles y similar al proceder humano de los afectos, dicha reacción sería programada, codificada, exenta de la espontaneidad de los gestos sinceros y de la libertad que mueve el verdadero altruismo y la entrega incondicional. No puede haber artificialidad en las emociones, a no ser que dejen de ser emociones.

Revalorización y pérdida del trabajo

En este planteamiento de lo que el robot hace o deja de hacer por las personas, cabe recuperar la afirmación con la que terminaba el capítulo anterior, donde se señalaba que los avances de la robótica y la inteligencia artificial, lejos de atemorizarnos ante la falsa perspectiva de que están muy cerca de desbancar a la humanidad, deben servirnos como espejo en el que contemplar nuestras singularidades, reconocer aquello que pertenece en exclusiva a la condición humana y resulta inalcanzable para la máquina, e identificar qué tareas se basan en esas capacidades.

En esta confrontación hay fracasos inevitables, pero también oportunidades. Entre estas últimas, la de potenciar aquellos trabajos que se sostienen sobre la base de competencias y habilidades singulares para la humanidad y son inimitables para las máquinas.

En las historias de Asimov las personas destacan, ante todo, por su creatividad y capacidad de liderazgo, por el ejercicio del pensamiento crítico, de la persuasión y la negociación, cuando no por sus intuiciones. Son las profesiones más valiosas –las menos útiles, pero las más valiosas de todas–, las que tienen su fundamento en estas destrezas –docentes, investigadores, trabajadores sociales y asistenciales, orientadores, artistas, entre otras–, las que no se pueden automatizar. Han tenido que venir las máquinas para ponerlas en el lugar que les corresponde.

Pero incluso si se pudieran automatizar, la pregunta que deberíamos hacernos es –siempre y cuando sea cierto que la humanidad es dueña de su destino tecnológico–: ¿queremos automatizarlas? ¿Por qué querríamos automatizarlas? Si alguien puede programar una máquina capaz de escribir una obra literaria a la altura de El Quijote de Cervantes, es porque esa misma persona puede escribir, de hecho, una obra literaria a la altura de El Quijote. ¿Qué sentido tiene entonces que sea una máquina y no la persona quien la escriba?

¿Queremos automatizar los servicios legales prestados por un abogado? ¿Hasta qué punto o instancia confiaríamos en el veredicto de la máquina? ¿Queremos automatizar el diagnóstico médico de un especialista? La consulta legal y médica ya se puede realizar en distintas plataformas de inteligencia artificial generativa –bien es cierto que en ocasiones se debe sortear la precaución impuesta por los desarrolladores–. Basta con cargar en la aplicación el texto objeto de disputa o la radiografía objeto de preocupación y plantear las preguntas u órdenes oportunas. Estos recursos pueden ser útiles como emergencia o aproximación, sin duda, pero ¿sentimos que tienen validez cómo veredicto final? ¿Confiamos para asuntos tan personales, íntimos y trascendentes en algo que no piensa, que no siente, y que por tanto no es capaz de empatizar con nosotros?

En ámbitos menos trascendentales –menos trascendentales para la vida humana en un cierto sentido, no en todos–, la traducción mediante inteligencia artificial está siendo atajada por el mundo editorial mediante una prohibición directa de su uso. Ya hay agentes literarios que no venden los derechos de traducción de los autores a su cargo si la editorial no garantiza que la traducción será llevada a cabo por una persona.

Aun con todas las prevenciones y esperanzas, la parte de las oportunidades y la ocasión de ensalzar la presencia humana y de las humanidades en el ámbito del trabajo no debe conducir a permanecer ciegos e insensibles ante el riesgo más inmediato y real: la cuestión del desempleo causado por la robótica y la inteligencia artificial, que ya se está haciendo notar e irá a más conforme mayor sea su desarrollo, y que se manifiesta al modo de «la historia se repite».

Volvemos a vivir un punto de inflexión en el mundo del trabajo, similar –sin caer en historicismos– al que pudo vivirse durante las grandes revoluciones industriales de los siglos XVIII y XIX. Entonces y ahora, las máquinas vuelven a sustituir a las personas en sus puestos de trabajo porque realizan con mayor exactitud, rapidez o eficiencia que nosotros la tarea que se nos ha encargado.

En este sentido, se podría decir con Asimov que «cualquier máquina es un robot. Una máquina de coser puede coser más deprisa que un ser humano, una taladradora puede penetrar una superficie dura

más rápidamente de como puede hacerlo un ser humano sin ayuda, un aparato de televisión puede detectar y organizar ondas de radio como nosotros no podemos hacerlo, y así sucesivamente» (VR, 13).

La diferencia en el momento actual de irrupción de la robótica es que los requerimientos crecientes de una demanda global y extremadamente competitiva hacen que la participación humana, siquiera la supervisión del trabajo de los robots, lastren la cadencia de una oferta exigente en cantidad y celeridad, que ha de incorporar transformaciones a una velocidad muy superior al ritmo de adaptación de la fuerza de trabajo humana. Además, la inversión en formación de nuevas competencias para los trabajadores humanos cada vez que se presenta un nuevo escenario de producción es muy grande en comparación con lo poco que cuesta reprogramar una máquina o crear nuevas copias de la misma adaptadas a los requerimientos de la nueva demanda, eliminando sin más las anteriores.

Habrá excepciones, por supuesto, sobre todo en aquellas profesiones menos industrializadas, donde el capital puramente humano siempre ha sido el centro de su desempeño. Por volver al ámbito de la traducción, el escritor y editor José Francisco Ruiz Casanova, doctor en Filología Española, ha señalado que la irrupción de la inteligencia artificial es de un modo u otro inevitable, si bien las máquinas no tienen por qué llegar a suplantar al traductor a todos los niveles, dado que sus carencias en la comprensión del contexto histórico, social y político del idioma, o del contexto vital del autor pueden hacer imposible una traducción mínimamente aceptable de ciertas expresiones y seguir haciendo necesaria la participación de aquel.

Sin embargo, incluso en este supuesto, los dos modelos de producción editorial más probables para la traducción en el futuro no son demasiado halagüeños para los traductores –sin mediar las prohibiciones expresas a la intervención de las máquinas que mencionábamos más arriba–:

1) una estructura editorial semejante a la actual, pero en la que los traductores humanos pasarían a formar parte de la

etapa última de edición, pues la versión del texto correría a cargo de la IA, esto es, una industria editorial puesta en manos de sus propias licencias de uso de la IA de traducción y en la que el trabajo humano, profesional, de edición seguiría existiendo, o

2) una estructura editorial en la que solo se precisarían versiones de lo publicado en su lengua original y una licencia de uso integrada en el libro electrónico que traduciría a nuestra lengua un texto que, en realidad, hemos adquirido en lengua inglesa, francesa, rumana, china o albanesa (32).

El caso, aun perteneciendo a un área de conocimiento donde el ser humano tiene todavía mucho que aportar al proceso de trabajo, nos hace ver que la inclusión de los robots en la cadena de producción tiene o tendrá como resultado, en un buen número de ocasiones, no tanto la reformulación o transformación del puesto de trabajo y de la tarea a llevar a cabo por las personas, como la expulsión directa de la mano de obra humana.

Estos momentos de transformación se pueden poner en perspectiva con el paso del tiempo, y acaban por resultarnos «naturales» o «inevitables» cuando ya se han afrontado y superado, pues es obvio que sin ellos no habríamos llegado donde estamos –sea el punto «donde estamos» mejor o peor, necesario o innecesario, ese es otro debate–. Sin embargo, esos mismos cambios suponen una auténtica conmoción para la sociedad que los vive y debe tomar las riendas del proceso de adaptación. En el momento de escribir estas líneas vivimos una de esas coyunturas de gran turbación social como resultado de las transformaciones que debemos afrontar en todos los órdenes –no solo el tecnológico–.

En el ámbito del trabajo, aun cuando los mismos avances de la tecnología y de la robótica rescaten y valoren antiguas profesiones artesanales y manufactureras dotadas de una autenticidad inigualable frente la artificiosidad de la máquina; aun cuando se revaloricen disciplinas del pensamiento antes descartadas por una

visión utilitarista de la realidad; aun cuando se creen nuevos perfiles profesionales y tipos de empleos para la creación y mantenimiento de estos sistemas tecnológicos, cabe preguntarse si esta creación de empleo será la suficiente en cantidad para paliar el factor humano o la fuerza de trabajo perdidos o que se perderán en los sectores todavía menos afectados, pero prestos a ser invadidos por la automatización de la robótica y la inteligencia artificial (puestos administrativos de nivel bajo y medio, transporte y carga de materiales, contabilidad, gestión de la información legal, atención al cliente virtual y física en grandes comercios).

En un plano meramente especulativo, también cabría preguntarse si esta «inutilidad» laboral, con la consiguiente falta de propósito para la vida de las personas afectadas por esta coyuntura y un reforzamiento de la mentalidad colectiva sobre «lo innecesario de ser tantos en el mundo», no servirá de base o palanca para un decrecimiento poblacional generalizado –derivado de las circunstancias, o más o menos intencionado–, o un refuerzo más –entre otros muchos de distinta índole que operan en ese mismo sentido– de la tendencia decreciente o desplome de los nacimientos –un desmoronamiento que, como tal, nada tiene de especulativo, pues ya existe y es plenamente identificable en países desarrollados–.

En ese potencial escenario cobraría fuerza el devastador panorama que dibuja Asimov en buena parte de su ficción. Mundos que ya hemos mencionado, en los que la vasta mayoría de la población humana ha sido suplantada por robots y el número de nacimientos es controlado al extremo –los que se van por los que vienen, en un estremecedor cálculo matemático de la vida–, pues no son necesarias más personas en el mundo: no hay trabajo para ellas, por lo que no tienen nada que aportar ni función alguna que desempeñar en el organismo social.

No cabe duda de que la cosmovisión desplegada en estos mundos espaciales y de ficción resulta altamente utilitarista y hace gala de un reduccionismo palpable en su ponderación de la valía de las personas, equivalente –reducida– al trabajo. Pero entonces cabría plantearse cómo de importante es para nosotros, ahora, en nuestro mundo de no ficción el trabajo, cuánto de nuestro día ocupa, si

facilita la formación de una familia y el tiempo de ocio que pasamos con quienes apreciamos o, por el contrario, el trabajo ha totalizado nuestra vida y el proceso de aislamiento del individuo y la comprensión de las relaciones empieza a pasar, de una forma u otra, por el trabajo y nada más que por el trabajo.

La entrada de la inteligencia artificial y de la robótica en el mundo laboral ha dado continuidad al empeño iniciado por Internet y las aplicaciones conectadas que, como el cliente de correo electrónico y los programas de gestión de equipos en entornos digitales –videollamadas, chats de grupo– entraron en escena con la promesa de facilitar y aliviar nuestra carga de trabajo. El hecho racional de estas tecnologías emergentes es hacer por nosotros el trabajo más tedioso y maquinal –valga la redundancia– y así ofrecernos la oportunidad de realizarnos en tareas más reflexivas y creativas.

En definitiva, dotarnos de más tiempo. Pero ¿tiempo para qué? Tiempo para trabajar más. No es raro que la manera más habitual de publicitar el aprendizaje de herramientas de inteligencia artificial tenga por bandera el mantra «incrementa tu productividad», ni que la inclusión de estos recursos en el trabajo allane el camino para la instauración del teletrabajo, que no es más que –de no serlo en la teoría, sí lo es en la práctica– otra manera engañosa de hacernos trabajar más.

Claro que ni tanto ni tan poco: si el resultado de la automatización va a ser, por el contrario, la instauración de sociedades lúdicas o de una «sociedad del ocio» libre de responsabilidades y abandonada a sí misma como tanto se quiso en las décadas de 1960 y 1970 –nunca ocurrió, aquí estamos– entonces conocemos bien nuestro destino: Solaria, Aurora, sociedad donde la humanidad dice vivir bien, pero en el fondo vaga sin rumbo.

En la búsqueda de un verdadero equilibrio, junto con Asimov resulta igualmente interesante y esclarecedora la advertencia de otro padre fundador –en este caso de la cibernética– algo anterior, pero contemporáneo del escritor. El también estadounidense Norbert Wiener supo ver las ventajas de la automatización –para el tiempo en el que escribió, en la producción en cadena–, pero junto con su defensa planteó los riesgos que acompañan a una mal entendida filosofía del progreso. La que se entiende no como un

incremento de nuestro desarrollo personal –cultural, espiritual, familiar– que revierte en un beneficio social mayor, sino como una vía de incremento del rendimiento y el beneficio. Esta última interpretación del progreso era la que Wiener veía como la más probable –el rumbo actual de los acontecimientos no ha cambiado la apuesta– (33), y el título del libro en el que se incluye su reflexión es todo un ruego y toda una declaración de intenciones: El uso humano de los seres humanos. En él no comparecen las máquinas, porque la situación generada de su inclusión en el puesto de trabajo no es culpa de ellas. Los culpables son, por volver a la propuesta de Asimov, quienes, contagiados de esa actitud economicista disfrazada de progreso social las crean con ese propósito, deciden implantarlas en el puesto de trabajo y, de otra parte, quienes las usan de ese modo sin oposición alguna.

Al respecto cabe señalar lo que Michel Servoz y Mark Coeckelbergh (34), especialistas de alto nivel en inteligencia artificial de la Comisión Europea, han advertido en múltiples informes y monografías: que se nos va el tiempo pensando o elucubrando lo que será de nosotros o lo que podrá pasar en un futuro a medio o largo plazo con la presencia creciente de la robótica y la inteligencia artificial en nuestras vidas, cuando el quid de la cuestión reside en decidir, porque tenemos la capacidad de hacerlo –dado que son nuestros semejantes quienes poseen la capacidad de desarrollarlas en uno u otro sentido–, qué debe pasar con la robótica y la inteligencia artificial. Somos nosotros quienes ponemos a los robots en el sitio que les corresponde.

De otra parte, la compensación por el trabajo decrecerá, o no se vaticina suficiente. Dado que muchos puestos de trabajo serán apoyados en sus tareas más duras por las máquinas, el trabajador no necesitará de una formación tan cualificada y su compensación será acorde con la misma, es decir, menor. Un mismo destino parecen correr las tareas de supervisión a las que muchos humanos serán destinados, y no solo en cuanto a la revisión a la baja de su remuneración, sino en cuanto a su carácter temporal: en cuanto el artefacto perfeccione sus funcionalidades, cualquier intervención humana dejará de tener sentido. Habrá puestos especializados en los que la presencia o manejo humanos serán o deberían ser necesarios; la cirugía, aun con la intervención de brazos robóticos

que potencian la precisión de las operaciones –el robot cirujano Da Vinci– seguirá necesitando de una toma de decisiones en tiempo real numerosa y grave por parte del especialista.

▪

29 Mayor, op. cit., pp. 121-148.

30 Sparrow, y Sparrow, «In the hands of machines? The future of aged care», Minds and Machines, 16, p. 18. Las cursivas son de los propios autores.

31 Nozick, Anarquía, Estado y utopía, p. 47.

32 Ruiz Casanova, ¿Sueñan los traductores con ovejas eléctricas?, p. 79.

33 Wiener, The Human Use of Human Beings: Cybernetics and Society, pp. 188-189.

34 Coeckelbergh, La ética de los robots.

Conclusión

En la narrativa de Asimov no siempre es necesario viajar a otros mundos espaciales, con sociedades ficticias sostenidas por una base cultural distinta a la nuestra, para reflexionar acerca de la tentación de caer en la pasividad más absoluta y la desactivación de nuestros anhelos bajo la errónea creencia de que los robots se bastan para acometer la empresa de la humanidad con los mismos propósito y espíritu.

En la Tierra, en un tiempo muy anterior a los años en los que transcurre la acción en las novelas de Elijah Baley y R. Daneel Olivaw, si bien un tiempo coincidente con nuestros días, pues según la cronología establecida en Yo, robot la robopsicóloga Susan Calvin vivió la mayor parte de su vida en la primera mitad del siglo XXI –se doctora y comienza a trabajar en U. S. Robots en el año 2008–, Asimov especula con situaciones muy cercanas a nuestra experiencia en las que cabría plantearse si un robot podría llevar a cabo ciertos trabajos decisivos para el devenir de una nación mejor que un ser humano.

Hacia el final del relato La evidencia, en el que la trama gira en torno a despejar la duda de si el prestigioso abogado y candidato a alcalde Stephen Byerley es o no un robot, Calvin, que en el fondo y aunque ha defendido su caso, tiene el convencimiento de que sí lo es, conversa con él y, siendo como es clara partidaria de las máquinas a las que ha dedicado toda su vida, expresa su preferencia por los robots para el desempeño de los cargos públicos:

—Me gustan los robots. Me gustan mucho más que los seres humanos. Si fuese posible crear un robot capaz de ser funcionario civil, creo que haríamos un gran bien, ya que las Leyes de la Robótica le impedirían dañar a un ser humano, lo incapacitarían

para la tiranía, la corrupción, la estupidez, el prejuicio. Y una vez hubiese servido durante un período prudencial, dimitiría, aunque fuese inmortal, porque sería incapaz de perjudicar a los seres humanos haciéndoles saber que habían sido gobernados por un robot. Sería el ideal.

–Excepto por el hecho de que un robot puede fallar, debido a la inherente inadaptación de su cerebro. El cerebro positrónico no tiene nunca la complejidad del cerebro humano.

–Tendría consejeros. Ni siquiera un cerebro humano es capaz de gobernar sin ayuda. (YR, 330)

Susan Calvin introduce en esta conversación un dilema sobre el que reflexionar y entablar un interesante debate. Es cierto, como indica la robopsicóloga, que con robots gobernantes salvaríamos, gracias a los condicionantes introducidos por las Tres Leyes, muchos de los defectos más habituales de nuestros políticos. No cabe duda de que los vicios y errores más inmediatos quedarían fuera de la ecuación. Sin embargo, la política es, en una mayoría de las ocasiones, no tanto la toma de decisiones técnicas para la resolución de problemas inmediatos como la proyección y materialización de una idea, no pocas veces a futuro y a largo plazo, sobre la mejor manera de aspirar al bien común. La abstracción de una cosmovisión sobre la vida buena para una sociedad –¿Qué es el bien? ¿Qué es el bienestar? ¿Qué es la libertad, la igualdad, la justicia?– insertada en los sistemas de computación de un robot o un súper ordenador dirigente requeriría de un consenso previo acerca de qué significa esa «vida buena».

En cualquier caso, y aun cuando se alcanzase un improbable consenso nacional o universal sobre tales cuestiones, difícilmente sería posible reducirlo a una fórmula única y cerrada, es decir, a los números y la fría lógica de un sistema de computación, aunque solo fuera porque necesitaría de factores para su aplicación –la íntima comprensión del sentir de cada tiempo y lugar, por ejemplo, o cierta dosis de esperanza, compasión o caridad– cuya contemplación y aprecio solo son factibles, cuando siente esa llamada y se presta, para el ser humano.

Y aun cuando podamos situarnos del lado de Susan Calvin y pensar que las máquinas son capaces de ejercer un buen gobierno, y que merece la pena confiar el futuro de nuestras sociedades a la perfección planificadora de los robots, todavía podríamos preguntarnos con Rodolfo Martínez: «¿Es preferible ser tutelados por un benévolo robot que no tiene otra prioridad que nuestro bienestar o somos lo bastante adultos para cuidar de nosotros mismos? Incluso, aunque no lo seamos, ¿no tenemos acaso derecho a ser los artífices de nuestro propio destino, aunque eso nos conduzca al desastre?» (35).

Volviendo al mundo de ficción espacial, a una conclusión parecida habrían llegado los propios robots según las leyendas atesoradas por Dom, una de las manifestaciones humanas de Gaia –el planeta de índole panteísta de Los límites de la Fundación y Fundación y Tierra–. En la línea narrativa alternativa abierta por esta leyenda, con la que Asimov ensaya una conexión tentativa y de corto recorrido entre las series literarias de los robots y de la Fundación con su extraordinaria novela El fin de la Eternidad –línea narrativa alternativa que no va más allá de la media docena de páginas, pero que resulta de sumo interés para la reflexión que hemos establecido aquí–, los robots habrían optado ellos mismos por su desaparición, cuando perciben que la interpretación cada vez más generosa de la Primera Ley y de su correspondiente papel de protectores de la humanidad, amplificada además por la incorporación de facultades telepáticas –lectura a discreción y sin ambages de la intencionalidad y el pensamiento de las personas–, les habría llevado a establecer un control férreo y limitante sobre la humanidad:

A medida que los robots fueron adquiriendo más inteligencia y versatilidad, interpretaron esas leyes [las Tres Leyes de la robótica de Asimov], en especial la primera, con creciente generosidad y asumieron, cada vez más, el papel de protectores de la humanidad. La protección ahogó a las personas y se hizo insoportable.

[...]

Con el tiempo, los robots progresaron hasta llegar a ser suficientemente humanos para comprender por qué los seres

humanos no querían que se les privara de todo lo humano con la excusa de su propio bien. A la larga, los robots se vieron obligados a admitir que quizá la humanidad se sentiría más a gusto cuidando de sí misma, aunque lo hiciera con negligencia e ineficacia. (LF, 447, 448)

La manera en que Asimov expresa este reconocimiento de los propios robots –llegaron a ser «suficientemente humanos para comprender por qué los seres humanos no querían que se les privara de todo lo humano»– es toda una declaración de intenciones.

En el desentendimiento e irresponsabilidad progresivos por los que han optado las personas, tienen que ser las máquinas, que ahora tienen y ejercen el control y podrían no devolverlo, las que, sin ser humanas, pero en la comprensión de cualidades eminentemente humanas como la voluntad y la libertad, den el paso de autoextinguirse. Eso es tener suerte.

Menos fortuna corren los habitantes de la Tierra de Vida y tiempos de Multivac, el relato culminante, más distópico e interesante de todos los publicados protagonizados por Multivac, una computadora que, tras las grandes catástrofes acontecidas en un ficticio siglo XXI, ha dirigido la reconstrucción de la civilización y en el proceso ha obtenido, de manera progresiva, más y más poder y control sobre todos los estamentos de la sociedad. Y como hemos observado en ejemplos anteriores, el protagonismo creciente de Multivac tiene lugar, no por ambición de la computadora, sino por la propia dejación de funciones, la docilidad en aras de la comodidad de la ciudadanía.

En los tiempos de las grandes catástrofes llegaron a perecer multitudes, y los ordenadores salvaron los restos y dirigieron la reconstrucción (y mejoraron su propio diseño, hasta que todos se fusionaron en Multivac), y los cinco millones de seres humanos que quedaban en la Tierra vivían con perfecta comodidad. (CC2, 645)

[Multivac] Era una presencia planetaria integrada por cables, fibras ópticas y microondas. Tenía un cerebro dividido en cien auxiliares, pero actuaba como uno. Existían terminales por doquier y ninguno de los cinco millones de seres humanos se encontraba lejos de alguno de ellos. (CC2, 648)

En el mundo dirigido por esta superinteligencia se puede hacer lo que se quiera, siempre y cuando lo que se quiera no atente contra la felicidad. Lo que constituye esa felicidad lo determina, por supuesto, Multivac: «Todos los hombres y las mujeres de la Tierra podían desarrollar un modo de vida según sus gustos personales, siempre que Multivac, que evaluaba cualquier asunto humano con perfecto criterio, no juzgase que ese modo de vida atentaba contra la felicidad humana» (CC2, 645).

El control ejercido por Multivac en el momento en el que se ambienta el relato es totalizador. En aras de la armonía y la paz se ha impuesto un régimen en el que gobierna la lógica más absoluta, el proceder logarítmico y un orden maquinal, un sistema que encorseta al ser humano, en el que no cabe el alma de la humanidad. Una esclavitud revestida de la piel de cordero de la seguridad.

La trama nos plantea que todavía queda un pequeño reducto de rebeldes, quince personas conocidas como «el Congreso». Una de sus miembros, Noreen, anima a uno de sus integrantes en apariencia más reacios a luchar y contrarrestar la situación, Bakst, a combatir el gobierno de Multivac y recuperar la verdadera libertad perdida.

—¿Lo has olvidado? ¿Todos lo habéis olvidado? ¿Recordáis otros tiempos? ¿Recordáis el siglo veinte? Ahora vivimos mucho tiempo, vivimos seguros, vivimos felices.

—Vivimos sin objetivos.

—¿Queréis volver al mundo tal como era antes?

Noreen negó con la cabeza.

—Fábulas para amedrentarnos. Hemos aprendido la lección. Con la ayuda de Multivac hemos salido adelante; pero ya no necesitamos esa ayuda. Si seguimos recibiéndola, nos ablandaremos hasta morir. Sin Multivac, nosotros dirigiremos los robots, nosotros dirigiremos las granjas, las minas y las fábricas.

—¿Con cuánta eficacia?

—La suficiente. Mejor aún con la práctica. Necesitamos ese estímulo, de todos modos, o moriremos.

—Tenemos nuestro trabajo, Noreen. El trabajo que escojamos.

—El que escojamos mientras no sea importante, y aun eso nos lo pueden arrebatar caprichosamente. (CC2, 647)

Como se deja entrever en la conversación, y se percibirá in crescendo conforme el relato avance hacia su desenlace final, la libertad no es un bien gratuito. La libertad viene acompañada del peaje de las responsabilidades que la humanidad –léase, no un ente abstracto, sino cada uno de nosotros– debe pagar a diario con gozo para disfrutarla. Un sacrificio y un contento que nos permiten seguir siendo humanos y vivir como tales.

▪

35 Martínez, op. cit., p. 84.

Lecturas de Asimov para este libro

Para iluminar los grandes debates sobre robótica einteligencia artificial seleccionados para este libro, se ha realizado una revisión sistemática de la obra de Asimov concerniente a los robots.

Teniendo siempre presente su accesibilidad para los lectores en las ediciones disponibles en español, el interesado en el tema, en la obra de Asimov o en la riqueza de los aportes del vuelo de la imaginación de la ciencia ficción, encontrará en los siguientes títulos estímulo y argumentos para la reflexión: el conjunto de relatos y ensayos breves Visiones de robot; la tetralogía del detective Elijah Baley y el robot Daneel Olivaw Bóvedas de acero, El sol desnudo, Los robots del amanecer y Robots e Imperio –aunque en esta última Elijah Baley ceda el protagonismo central a otros personajes–; y la bilogía final de la saga Fundación, compuesta por las novelas Los límites de la Fundación y Fundación y Tierra, en las que dicha saga queda vinculada por el desarrollo de su trama al ciclo de los robots, fundiendo ambas series literarias en una. Estas lecturas componen el grueso de la obra utilizada para aportar respuestas a las preguntas formuladas a lo largo de este libro acerca de la robótica y la inteligencia artificial.

Como complementos a esas lecturas principales, se suman otras que aportan detalle y concreción: el fix-up de Yo, robot –relatos en su mayoría ya presentes en Visiones de robot, si bien con presencia de algún título extra y unidos por el hilo conductor de la célebre robopsicóloga Susan Calvin–; los relatos sobre robots o sobre la computadora Multivac presentes en los dos volúmenes en español de Cuentos completos; y las tres novelas –del total de seis– pertenecientes a la saga de Lucky Starr –con un enfoque mucho más juvenil– en las que hacen aparición los robots, El gran sol de Mercurio, Las lunas de Júpiter y Los anillos de Saturno.

El conjunto narrativo utilizado para este libro no es todo lo

publicado sobre robótica e inteligencia artificial por Asimov, pero sí la mayor parte y, en el caso de las «lecturas principales», de lo más leído y conocido del autor y, sobre todo, lo más interesante desde la perspectiva de las ideas sobre el tema tratado.

En el caso de aquellos lectores que no hayan abordado todavía la obra sobre robótica e inteligencia artificial de Asimov, y quieran hacerlo con cierta garantía de orden y concierto, propongo seguir el listado de lecturas proporcionado a continuación, en el que se ha sumado, después de Robots e Imperio, la trilogía inicial de la saga Fundación. Esta trilogía no contienen referencia alguna a los robots, pero su lectura es necesaria para preparar y entender Los límites de la Fundación y Fundación y Tierra, quedando el total de lecturas y su orden de la siguiente manera: Visiones de robot, Bóvedas de acero, El sol desnudo, Los robots del amanecer, Robots e Imperio –hasta aquí el ciclo de los robots–, Fundación, Fundación e Imperio, Segunda Fundación, Los límites de la Fundación y Fundación y Tierra –hasta aquí las cinco de las siete novelas de la saga Fundación que terminan por relacionarse y cierran el círculo con el ciclo de los robots–.

En cuanto a la bibliografía básica escogida para trabajar y citar en este ensayo, me he decantado por aquellas ediciones más accesibles para una mayoría de lectores, por su precio y por su disponibilidad en una amplia variedad de formatos de lectura. Suponen la excepción las novelas de Lucky Starr, pues la edición con la que se ha trabajado es antigua y está descatalogada, si bien, como lecturas complementarias que son, se mencionan y citan en pocas ocasiones en este texto y, en cualquier caso, pueden encontrarse todas en ediciones actuales.

Bibliografía

ASARO, P. M.: «What Should We Want From a Robot Ethic?», International Review of Information Ethics, 6, 2006.

ASIMOV, Isaac: Lucky Starr. El gran sol de Mercurio, Bruguera, Barcelona, 1976. (GSM)

——: Lucky Starr. Los anillos de Saturno, Bruguera, Barcelona, 1977. (AS)

——: Lucky Starr. Las lunas de Júpiter, Bruguera, Barcelona, 1977. (LJ)

——: Lucky Starr. El ránger del espacio, Bruguera, Barcelona, 1976.

——: Lucky Starr. Los piratas de los Asteroides, Bruguera, Barcelona, 1976.

——: Lucky Starr. Los océanos de Venus, Bruguera, Barcelona, 1977.

——: Los límites de la Fundación, Debolsillo, Barcelona, 2015. (LF)

——: Fundación y Tierra, Debolsillo, Barcelona, 2016. (FT)

——: Cuentos completos 1, Debolsillo, Barcelona, 2019. (CC1)

——: Cuentos completos 2, Debolsillo, Barcelona, 2019. (CC2)

——: Visiones de robot, Debolsillo, Barcelona, 2021. (VR)

——: Yo, robot, Edhasa, Barcelona, 2022. (YR)

——: Bóvedas de acero, Debolsillo, Barcelona, 2022. (BA)

——: El sol desnudo, Debolsillo, Barcelona, 2022. (SD)

——: Los robots del amanecer, Debolsillo, Barcelona, 2022. (RA)

——: Robots e Imperio, Debolsillo, Barcelona, 2022. (RI)

——: Trilogía Fundación, Nova, Barcelona, 2022.

——: Yo, Asimov. Memorias, Arpa, Barcelona, 2023.

BODEN, M. A.: Inteligencia artificial, Turner, Madrid, 2017.

BOSTROM, N.: Superintelligence. Paths, Dangers, Strategies, Oxford University Press, 2016.

COECKELBERGH, M.: La ética de los robots, Cátedra, Madrid, 2024.

HEIDER, F.: The Psychology of Interpersonal Relations, Martino Fine Books, 2015.

HERMANOS ČAPEK: R. U. R. y El juego de los insectos, Alianza Editorial, Madrid, 1966.

JOHNSON, P.: Intelectuales, Javier Vergara Editor, Barcelona, 1990.

KAPLAN, J.: Inteligencia artificial. Lo que todo el mundo debe saber, TEELL, 2017.

KURZWEIL, Ray: La singularidad está cerca, Lola Books, Berlín, 2021.

LARSON, E. J.: El mito de la inteligencia artificial, Shackleton, Barcelona, 2023.

MARTÍNEZ, R.: La ciencia ficción de Isaac Asimov, Sportula, 2021.

MAYOR, A.: Dioses y robots. Mitos, máquinas y sueños tecnológicos en la Antigüedad, Desperta Ferro, 2019.

MOLINA, C. A.: ¿Qué hacemos con los humanos?, Deusto, 2023.

NOZICK, R.: Anarquía, Estado y utopía, INNISFREE, 2018, p. 47.

PEARL, J. y MACKENZIE, D.: El libro del porqué: la nueva ciencia

de la causa y el efecto, Pasado & Presente, 2020.

RUIZ CASANOVA, J. F.: ¿Sueñan los traductores con ovejas eléctricas?, Cátedra, Madrid, 2013.

SANDEL, M. J.: Justicia. ¿Hacemos lo que debemos?, Debate, Barcelona, 2011.

SPARROW, R. y SPARROW, L.: «In the hands of machines? The future of aged care», Minds and Machines, 16, 2006, pp. 141-161.

WIENER, N.: The Human Use of Human Beings: Cybernetics and Society, Houghton Mifflin, 1950.